

Mapmaking for Language Documentation and Description

Lauren Gawne

SOAS, University of London

Hiram Ring

Nanyang Technological University

This paper introduces readers to mapmaking as part of language documentation. We discuss some of the benefits and ethical challenges in producing good maps, drawing on linguistic geography and GIS literature. We then describe current tools and practices that are useful when creating maps of linguistic data, particularly using locations of field sites to identify language areas/boundaries. We demonstrate a basic workflow that uses CartoDB, before demonstrating a more complex workflow involving Google Maps and TileMill. We also discuss presentation and archiving of mapping products. The majority of the tools identified and used are open source or free to use.

1. Introduction¹ Linguists engage in representing linguistic data so that others can easily see and understand what a linguist is attempting to describe. There are many ways to represent a language and its context. Some are non-spatial, such as indexes and tables. Others are spatial, although the space may be abstract, such as in Neighbornets and Bayesian phylogeography. Representing the distribution of a language in geographical space is a useful visualization, and being able to create maps is a valuable skill. People are used to processing geographical and spatial cues on a map, and well-designed maps can help researchers convey facts about a language to a wider audience (Upton 2010). However, language maps such as those found in high school atlases can give the erroneous impression of a clear relationship between nation-states and homogeneous national languages (Mackey 1988:24–25), while even detailed language maps of some linguistically-diverse areas can obscure small languages (Dahl & Veselinova 2005). The majority of the world's languages are under-documented, and thus our knowledge of where these languages are located is also lacking. Often the

¹Lauren Gawne would like to thank the Research Bazaar team at The University of Melbourne for their TileMill training course as part of ArtsHack, with additional thanks to Steven Bennett for server space and answering endless questions. Lauren would also like to thank Christine Gawne, who did the tedious work of tracing the maps (duplicated in Figures 1 and 2) that Lauren used before she learned to make maps for herself. Thanks also to the CartoDB support team, and the makers of GoogleMaps and TileMill.

Hiram Ring would like to thank the programmers of GoogleMaps, CartoDB, and TileMill for creating excellent tools, and NASA for GPS. Hiram would also like to thank the Ethnologue team and members of SIL for introducing him to language mapping and the identification of small languages in remote and inaccessible places.



location of language-speaking populations is known only to the speakers themselves and possibly specialist researchers (Dahl & Veselinova 2005). The ability to make maps has also been limited to those with access to the technical skills and capital, creating an imbalance in who can claim the right to be represented (Peeters 1992:7).

In this paper we explore the “cartographic representation of linguistic facts,” to borrow a useful definition of linguistic geography from Kehrein et al. (2010:xi). This may refer to ‘language maps’ that map the location of one or more language varieties, or ‘linguistic’ maps that map particular linguistic features of one or more language varieties (Girnth 2010: 100). We particularly focus on languages that are generally underrepresented in the geographical tradition. As noted above, many maps relating to languages or dialects are out of date or based on incomplete information. This is particularly the case in remote or inaccessible areas where little is known about a particular language group or variety.

This paper introduces readers to language mapmaking in relation to the language documentation process. This includes some of the considerations necessary for making language maps, and the introduction of techniques and methods that have made the task of map creation much easier in the last few years (see Dahl & Veselinova 2005, who a decade ago observed a lack of an easily accessible tool for language mapping). These techniques are centered around making use of Global Positioning System (GPS) data with free software such as Google Earth² and TileMill³ and online services such as CartoDB,⁴ converting the data into features to overlay on a digital map of the world. We first discuss some considerations regarding mapmaking, particularly in the language documentation context (§2). We then identify ways of gathering GPS data (§3.1), how to convert the data into a usable format (§3.2), how to use the data with various free mapping software (§3.3–3.5), and how to share your map, including exporting the map as a print-ready image (§3.6).

Mapping has a long history in linguistic research, even if minority and endangered languages and varieties are generally underrepresented in the literature. Some of the earliest mapping of linguistic information by geographical area in the Western linguistic tradition includes the *Sprachatlas des Deutschen Reiches* by Georg Wenker and Ferdinand Wrede in 1888 (Scheuringer 2010) and the *Atlas Linguistique de la France*⁵ by Jules Gilliéron in 1902 (Swiggers 2010). Since then, maps have been used by linguists for a variety of purposes, such as the historical mapping of dialects (Moore et al. 1935), sociolinguistic variation and dialect studies (Trudgill 1983; Labov et al. 2006; Kretzschmar 2013b), language attitudes (Waller 2009), perception of variation (Alfaraz 2014; Bounds 2015; Yan 2015), as well as variation and change over time (Kulkarni et al. 2014; MacKenzie 2014; Prichard 2014), and many other areas of linguistics.

Mapping is not an entirely objective activity, and methods of mapping draw on the same theoretical underpinnings as any other tool of linguistic analysis. For example,

²<http://earth.google.com>.

³<http://www.mapbox.com/tilemill>.

⁴<http://cartodb.com/>.

⁵Online at: <http://diglib.uibk.ac.at/ulbtirol/content/titleinfo/149029>.

as Kretzschmar (2013a:53) observes, mapping language variation is grounded in a Neogrammarian assumption that there are ‘dialects’ which may exist and can be mapped (see also Labov et al. 2006:5; Murray 2010). While it is possible to undertake a kind of word geography, where points are plotted without any claim to dialects, variationist approaches may result in striving for neat generalizations in messy data (Kretzschmar 2013a:53–54). Mapping has also been a valuable tool in linguistic typology (Mühlhäusler 2010), and many of the mapping skills in this paper can also be applied to maps illustrating typological features.⁶

Unlike sociolinguists and others who work with relatively better-described languages, those who document languages are often providing a baseline of knowledge that was previously unavailable and which can serve as a stepping-stone for future research (cf. Roth 2013, building on previous work by SIL). Although we touch on some other uses of maps in linguistics, our aim here is not to produce a comprehensive discussion of the history of mapmaking in linguistics. Auer & Schmidt (2010) and Lameli et al. (2010) provide comprehensive discussion of the history and contemporary state of linguistic mapping, and further useful information can be found in chapters of Krug & Schluter (2013), as well as Kretzschmar et al. (2014).⁷ The *Journal of Linguistic Geography*⁸ was launched in 2013, and provides recognition that mapping is not simply a factual representation, but a core element of the “...pursuit of a better understanding of the nature of language structure and language change” (Labov & Preston 2013:1). The process we describe can be useful to linguists in many subfields, not just language documentation. We hope to encourage more accessible and engaging mapping and visualizations that can add to our knowledge and understanding of language and of human interaction.

Throughout this paper we will be drawing on maps that illustrate some facet of information about language, or about communities of language speakers. This will include maps of language groups that we have worked with. Author Gawne works with varieties of Yolmo (Tibeto-Burman, ISO 639-3 scp) that are spoken in Nepal. The traditional area that Yolmo is spoken in centers on the Melamchi and Helambu valleys north of Kathmandu; however, there are multi-village communities that have been living in other parts of Nepal for over a century (see Gawne 2013b for more detail). These include communities in Lamjung, Ilam, and Ramechhap (where the language is known as Kagate, ISO 639-3 syw). Gawne has worked mostly with the Lamjung Yolmo and Kagate varieties. Author Ring works with speakers of Pnar (ISO 639-3 pbv) and other Khasian languages spoken in Meghalaya state of North-East India, Austroasiatic languages in Thailand, and languages in North-East India which border Pnar areas, such as Biata (Tibeto-Burman, ISO 639-3 biu) and Assamese (Indo-Aryan, ISO 639-3 asm).

⁶WALS, mentioned in §2.1 is an example of this on a large scale. For an example of mapping a specific typological feature, see this map by the first author plotting the distribution of reported speech evidentiality in Tibeto-Burman languages: http://lgawne.cartodb.com/viz/95fc31aa-3543-11e5-bc55-0e0c41326911/public_map.

⁷A useful and accessible blog post from Hugh Paterson can be found here: <http://hugh.thejourneyler.org/2012/types-of-linguistic-maps-the-mapping-of-linguistic-features/>.

⁸http://journals.cambridge.org/jid_JLG

2. Considerations regarding mapmaking In this section we will discuss the nature of linguistic mapmaking, and considerations that will need to be addressed before starting the mapmaking process. This includes both practical and theoretical considerations. Even though language documentation often involves site-specific communities who have lived in an area for a long period of time, there has been a general lack of discussion regarding mapmaking in the language documentation literature. Bower (2008:111) mentions the need to take a device on fieldwork to collect GPS data and, where possible, use existing high-quality maps to mark features. The ability to generate your own high-quality maps from the GPS data collected has simplified this process somewhat, as we will show below.

Mapping involves the manipulation of what are actually a limited number of data types. In §3 we introduce file type conventions and manipulation, but it is worth covering some basic terminology here: points, lines, polygons, and layers.

Point data is the coordination of latitude and longitude represented as a single point. What is represented by a point depends on what is being mapped: it may be a physical object like a house, or a proxy object like mapping a house to represent the location that a person lives in, even though the person obviously moves from that point.

Line data is a series of two or more points that are joined. They can be used to represent physical information, such as paths or rivers, non-physical socio-political boundaries, potential dialect boundaries or migration pathways. Lines are also known as vectors in some of the literature.

Polygon data is a series of points that form a bounded shape. As with the other two data types, this may be a representation of physical object like a lake, or a more abstract bounded locale like a dialect region.

Layers are a way of organizing and styling map data. When thinking of maps it is best to think in terms of layers. Each thematically linked group of data of the same theme that is kept in the same file will be represented and manipulated in a map as a single layer. A layer will contain only the same type of data; that is, it will contain exclusively points, lines or polygons. Knowing whether to map points, lines or polygons is part of deciding what story you want your map to tell. We discuss some of the issues around this in §2.4.

Effective mapmaking skills allow researchers to make maps that specifically convey relevant information. By way of demonstration of the advantages of well-made maps, Figure 1 and Figure 2 are the maps used by Gawne in her completed PhD dissertation. This outline of Nepal was traced from an atlas, and then MS Paint was used to label the relevant language varieties. The villages in Figure 2 were identified in Google Earth using satellite imagery and the road traced in approximation. These maps offer limited contextual information and are limited in their reproducibility.

We are unable to reproduce the map Gawne was using earlier, as it was an image with copyright restrictions. Copyright-restricted lack of reproducibility is, in itself, sufficient reason for such a map to be unsatisfactory, but it also contained a great deal of irrelevant information and was not very visually appealing. In contrast to the hand-drawn and difficult to replicate maps above, Figure 3 is a map that Gawne

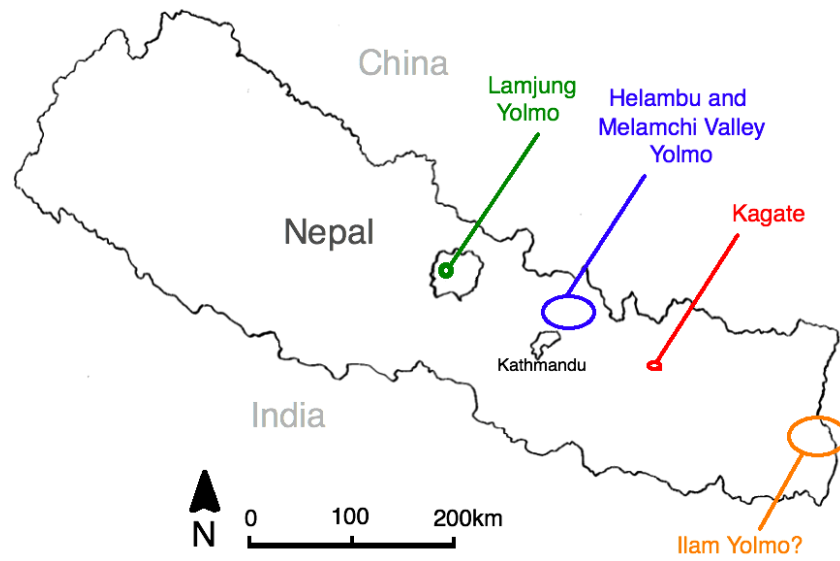


Figure 1. Map of Nepal with Yolmo villages used in Gawne (2013a)

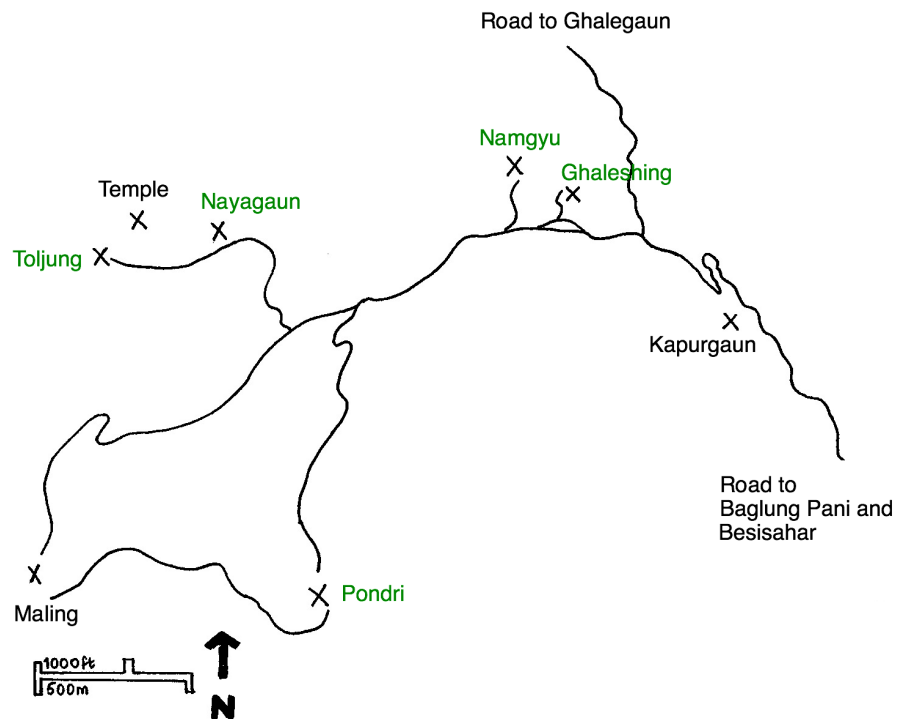


Figure 2. Map of Yolmo villages in Lamjung used in Gawne (2013a)

created after attending a mapping workshop, using TileMill and GPS data she collected. Unlike Figure 1 it is easily re-purposed and replicated, and looks much more professional. This map has appeared in subsequent publications including Gawne (2013b). In §3.5 we will demonstrate how to create a map like the one in Figure 3.

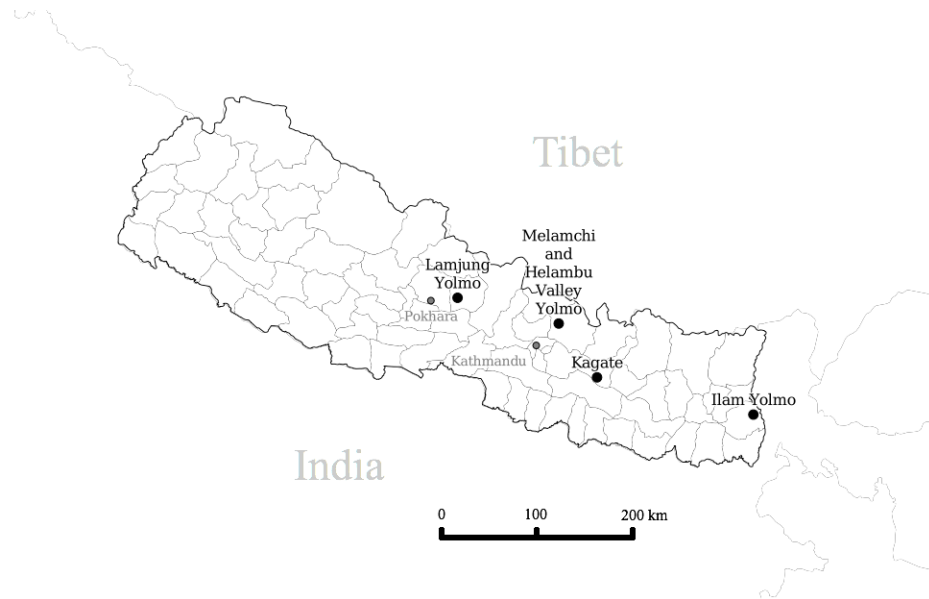


Figure 3. Map of Nepal made in TileMill used in Gawne (2013b)

2.1 Different types of maps In this section we introduce some further terminology to distinguish different types of mapping products that we will use throughout the paper. We identify a number of different map product types, such as paper maps and digital maps. Within digital maps we distinguish between static maps, dynamic maps, and multimedia maps. Each of these map types have their own specific advantages and limitations.

Print maps have a number of uses. Such maps are still required in traditional publishing formats. These maps are not dependent on digital literacy or access to an internet connection. Additionally, communities may appreciate having a physical copy of a map as a sign of legitimation (Bower 2008, ch. 14). Paper maps are often a product of digital maps, but may require different colors or designs to look good. Historically, print maps were created manually; Kretzschmar (1996:15) notes that the maps for the Linguistic Atlas of the Middle and South Atlantic States (LAMSAS) were originally all hand-drawn, taking up to six to eight hours to complete a single map of a single feature of lexical variation. Only then could isoglosses be observed, if they existed, and analyzed. This meant that the researchers had to be highly selective about which features to map.

Digital maps are those with recoverable geolocation information. It is possible to design a map using digital illustration software by drawing directly onto an existing

map. These images do not involve spatial encoding and so the data is not recoverable, and cannot be used to make other maps. Kretzschmar (1996) illustrates the benefits of keeping geospatial data separate from the image rendered. The same LAMSAS datasets that would take a full working day to draw were digitally encoded and in 1992 a GIS program took around 90 seconds to plot all the points. This meant that more maps could be generated in less time, allowing for quicker observation of potential isoglosses and full use of the survey data. As Labov & Preston (2013:1) observe, digital maps are not constrained by printed space, budgets for color printing, or other practicalities that have been holding linguistic geography back as a discipline. Not all digital maps are the same, however, and below we describe three varieties of digital map in the paragraphs that will be relevant to our discussion.

Static maps have a fixed number of features that are presented. Print maps are inherently static, as they are fixed. Digital maps may be static, although even in static digital maps it may be possible to scroll and zoom. In the WALS maps it is possible to also change the base map and the style of the points used, but the map still shows fundamentally the same data. Static maps can be made in all digital mapping software.

Dynamic maps, in contrast to static maps, allow the user to alter what information they see, be that through an ability to toggle certain features on or off or by showing different information at different zoom levels. The kinds of features that can be built into a dynamic map depend on the software used to build the map. For example, CartoDB gives you the option of publishing a digital map where the viewer can easily toggle features on or off. However, to build a map where different features appear at a certain level of zoom requires the user to move beyond the simple interface and to code that information using their knowledge of HTML/CSS. Dynamic maps can also represent changes and movement over time, which can be useful for illustrating migration patterns. Although easily created animated maps still have a way to go, Hanewinkel & Losang (2010) provide an introduction to the topic.

Multimedia maps integrate other media into a map. For example, hovering over or clicking on a village name may bring up an image of that village, or a recording of the village name. The Algonquian Linguistic Atlas⁹ has recordings linked to mapped points. Media can be linked to static maps, or form part of a dynamic map. Different software can allow users to build in multimedia links with varying degrees of ease.

The status of maps as objects has changed drastically in recent decades. Peeters (1992:7) and Dahl & Veselinova (2005) note that static maps must, in their nature, exclude and overcrowd information. While all maps are selective snapshots of a moment in time, modern digital mapping offers some solutions to the problems inherent in static maps.

Digital maps allow collected data points to be represented in different ways, so that multiple maps can be made to demonstrate the complexity of a linguistic situation. Also, each map feature set is produced on its own layer of the map, so additional information can be easily toggled on or off. Digital maps can include a zooming feature, allowing for detailed village-by-village views on closer zoom and

⁹<http://www.atlas-ling.ca/>

more generalized polygons at greater distance. These digital maps are obviously not suitable for print publications, and there are still the same concerns about what is included and excluded; however, they allow for more complex language stories to be told.

2.2 Currently available maps Linguists generally use one of a number of strategies when making language maps for presentation and publication. Some may use existing resources like Google Maps, dropping digital pins to locate languages, or may use image editing software to add that information to an existing static map. While this serves the necessary functional purpose, it is limited. Firstly, the information drawn onto the map is hard to re-purpose into another map, unlike digital mapping where GPS data points are created and stored separately from the styled map. Secondly, every map is designed with a specific purpose in mind (cf. MacEachren 1995), so co-opting an existing map may mean including extraneous information. Thirdly, there may be limits to reproducing such maps in publications for reasons of copyright and visual clarity (i.e., graphic resolution). Finally, these maps are often not particularly aesthetically appealing. Speakers of endangered languages are often minorities, rendered invisible on official maps. Making targeted and professional-quality maps for these languages can help increase their perceived status, both within the community and among the general public.

The most comprehensive language maps that currently exist are those created by the Summer Institute of Linguistics (SIL), which are updated periodically for editions of the *Ethnologue* (Lewis et al. 2013), based on the World Language Mapping System (WLMS).¹⁰ This information is also used to provide the GPS information for ISO 639-3 labelling,¹¹ which is used by services such as Glottolog¹² and OLAC.¹³ These maps can be useful, especially the ISO 639-3 point data; however, they also have some limitations. The first is that individual linguists and specialists would often prefer to have a map that illustrates different features than what these larger datasets offer. Linguists (and the larger linguistic community) would benefit from a way to create new maps based on their research and knowledge of a particular area, using GPS data to clarify and assist their efforts. These maps could then be posted online or be printed in publications. The second limitation is that these maps are licensed products of WLMS and cost money to obtain. Current publicly available software is free to use, and builds upon open access data. This means that maps created by researchers can be easily distributed online and in print without the licensing limitations that pre-made maps and map data can have.

There are a number of other projects that have a strong geo-spatial element and a focus on global linguistic diversity. Many researchers interested in typology will have used the World Atlas of Linguistic Structures (WALS).¹⁴ The discussion of each

¹⁰<http://www.worldgeodatasets.com/language/>.

¹¹<http://www.ethnologue.com/about/language-maps>

¹²<http://glottolog.org/>.

¹³<http://www.language-archives.org/REC/discourse.html>.

¹⁴<http://wals.info/>.

typological feature in WALS includes a map of the distribution of that feature across the languages surveyed. On a similar scale, the Endangered Languages Project¹⁵ employs a map based on Google Maps technology. The data that formed the basis of the initial launch of the project was put together by the Catalogue of Endangered Languages (ELCat) project, and is drawn from a range of sources. Each location source is cited on the language page. For example, the geolocation information for Yolmo (also known as Helambu Sherpa) is from Glottolog,¹⁶ while the geolocation information for Alyawarr and many other Australian languages comes from Bower (2011).¹⁷ If there is more than one source with differing geolocational information, the Endangered Languages Project cites both. Language Landscape¹⁸ is a web-based participation project that takes a different perspective on the distribution of language, and instead plots the geospatial location of an individual when they make a specific recording. This moves beyond the representation of languages as grounded in specific locations, and instead maps specific moments of language use.

There is also an increasing number of projects that effectively use language mapping for a specific language, or group of languages. The Algonquian Linguistic Atlas has a large collection of phrases spoken in Algonquian languages from across Canada; the viewer can click on each one to hear how the phrase is spoken in the language of that area. The Yup'ik Environmental Knowledge Project¹⁹ is also concerned with mapping indigenous language, but focuses on place names and other geographic knowledge of the Yup'ik communities. Bower's work on the history of Australian languages includes the production of mapping data as both a research tool and a research output.²⁰ There are many other useful examples of maps used for research and educational purposes. We hope the examples discussed here provide some illustration for the interested linguist about different ways in which it is possible to use language mapping to support both research and community-oriented purposes.

2.3 Ethical considerations Language maps can be problematic constructions. As static and selective representations of complex phenomena, they can act as both representations of power and tools of power (Luebbering 2013:49). As Peeters (1992:7) notes, there is a certain authority given to a place or a people by putting them on a map. For endangered language communities this may be a good thing, giving them recognition and status that they may never previously have experienced, even if those maps are only used within the community and research publications. But mapping can also be used as a tool of division, or evidence in territorial claims. Therefore, linguists need to think very hard about the potentially exclusionary or over-expansive story their map might tell.

¹⁵<http://www.endangeredlanguages.com>.

¹⁶<http://www.endangeredlanguages.com/lang/5710>. (The Glottolog data appears to be based on the Ethnologue information, and is available here: <http://glottolog.org/meta/downloads>)

¹⁷<http://www.endangeredlanguages.com/lang/5249>.

¹⁸<http://languagelandscape.org/>.

¹⁹<http://eloka-arctic.org/communities/yupik/atlas/index.html>.

²⁰<http://www.pamanyungan.net/>.

It is also worth considering whether the community will want their location mapped and distributed. While some groups may approve of this activity, it should not be assumed that everyone wants to be ‘on the map.’ It may be appropriate to only indicate their location very generally within the larger nation or state where they reside, and then present a more detailed map that doesn’t necessarily give traceable GPS information. For some communities, being locatable may be politically or socially undesirable. It may be appropriate to still collect geospatial data for the researcher’s project, but not make publicly available maps. There may also be contexts where it is not appropriate to collect this data, or to map anything other than a very generalized location.

Researchers should check with their IRBs about the handling, storage, and dissemination of geolocation data in relation to the specific context of each research project. We believe, however, that it is important to ensure that you do have a process in place for seeking permission from the community to use geospatial data. This is particularly the case if multimedia maps are made featuring images of community spaces linked to their specific geographical location. The kind of permissions you will need to ensure are contingent upon the social structure of the community you are working with; in some communities there may be a specific official body that it is best to approach about collecting this type of data (for more on this topic see Bown 2008:3; Chelliah & De Reuse 2011:124).

It is also worth considering whether all local geographic features are appropriate for mapping. Some communities may have sacred spaces or totem lands that they do not wish to have located. Including the community in any GPS coordinate plotting and sharing early map drafts with them can help ensure that they are comfortable with the amount of information you are building into a map. As in all language documentation data collection, sensitivity to context is key.

2.4 What to put on the map Choosing what to put on a map can change how it is read. Girth (2010) suggests asking four basic questions before starting a map:

1. What is the aim of the map?
2. What is mapped?
3. How is it mapped?
4. What mental picture does the map produce?

In this paper, we take it that the aim of the map is to represent the location of one or more languages or language communities. But for the second question, there may be other geological or political features that need to be considered. You may choose to include national or local borders to indicate the relationship between a language group and larger geo-political constructions. Including other language groups or urban centers may help contextualize the linguistic landscape. Some features of the local geography may provide specific challenges to map clarity that will need to be addressed; as one reviewer noted, the Himalayas provide a particular challenge in

this regard due to the rapid changes in elevation. Different map types will allow you to focus on different features and help you and the community visualize what is most important.

Within the target group being mapped, you may choose to map them as a single unit or break the groups down further. You will need to decide on what grounds you wish to make such distinctions and what divisions that may indicate in a map. Some groups that you have good reason to consider as separate dialects or languages on linguistic grounds may have strong social or cultural reasons to be identified as a single group.

Of course, the location of language groups may not be the only language-related mapping target. As Lubbering (2013:44) notes, there is a diversity of topics and variables, and their geographic arrangement, when it comes to languages. With the tools we demonstrate in this paper, it would also be possible to create maps that illustrate lexical variation between groups, density of language maintenance in particular areas, or areal contact.

Whatever you plan to map, considering the design elements will be vital to ensuring the map tells the story you want. This addresses Girnth's third question above. Ormeling (2010) provides a good overview of the kinds of elements you will need. Some will be more relevant to particular mapping projects than others.

The 'projection' is the way that the Earth's sphere is rendered in flat space. Some projections can make the polar areas appear relatively large compared to equatorial areas; however, if you are mapping a space the size of Germany or smaller, as is most likely when mapping smaller languages, any distortion from different projections should not be noticeable (Ormeling 2010:29). The 'zoom level' will determine how much terrain is visible in the mapped pane, be it the entire globe or a small pocket of land. The colors and forms you use for different map elements will also be part of the story. Points of the same size will be perceived to represent a similar number of speakers, and darker colors will be more prominent on the map and should be used to map key elements. Digital mapping allows you to try different design styles, and modifying different features will give you insight into what works best for the map you wish to make.

Accessibility for people with color-blindness is another design consideration to keep in mind, as some color distinctions are difficult to see for up to eight percent of the population. If you are using a red and green distinction it is possible this may not be visible to some members of your audience. There are a number of color-blindness simulators online so that you can see the effect of some design choices on common color-blindness types.²¹ The advantage of making digital maps is that it is very easy to modify color schemes, or if you are required to work with a particular palette, to also duplicate the map in color-safe contrasts as well.²²

Each language context will present its own challenges and points of interest for the task of mapping. Any map is inherently controversial, as no one map will satisfy

²¹We like: <http://www.color-blindness.com/coblis-color-blindness-simulator/>.

²²Duplication was the solution for these Australian variation maps from the Linguistics Roadshow, with a series of color safe maps: <http://lingroadshow.com/resources/englishes-in-australia/vocabulary/mapping-words-around-australia/mapping-words-around-australia-cs/>.

all users (Peeters 1992:6), but there are some generalizable considerations to be made. For this paper we are assuming that the map will be of the location of settlements in which the target variety is spoken, be they individual homesteads, small villages, or large towns and cities. You will have to decide if you want to mark all of these as individual points (which may be preferable if there are a manageable number of tokens), or to indicate the language area with a single polygon (an enclosed shape made from a collection of points; see GPS section below for more detail). You may wish to represent each village as a separate point, but then have larger polygons representing the influence of different contact languages or related languages. The following images from Google Earth demonstrate these differences in visualization: Figure 4 provides a map of some Pnar-speaking towns in NE India as individual points, Figure 5 represents these towns as a single polygon, and Figure 6 represents both points and polygon in relation to other polygons representing other language communities.²³ Each allows for a different understanding of the relationship between Pnar speakers and neighboring linguistic communities, giving different mental pictures of the distribution of Pnar as far as Girth's fourth question is concerned.

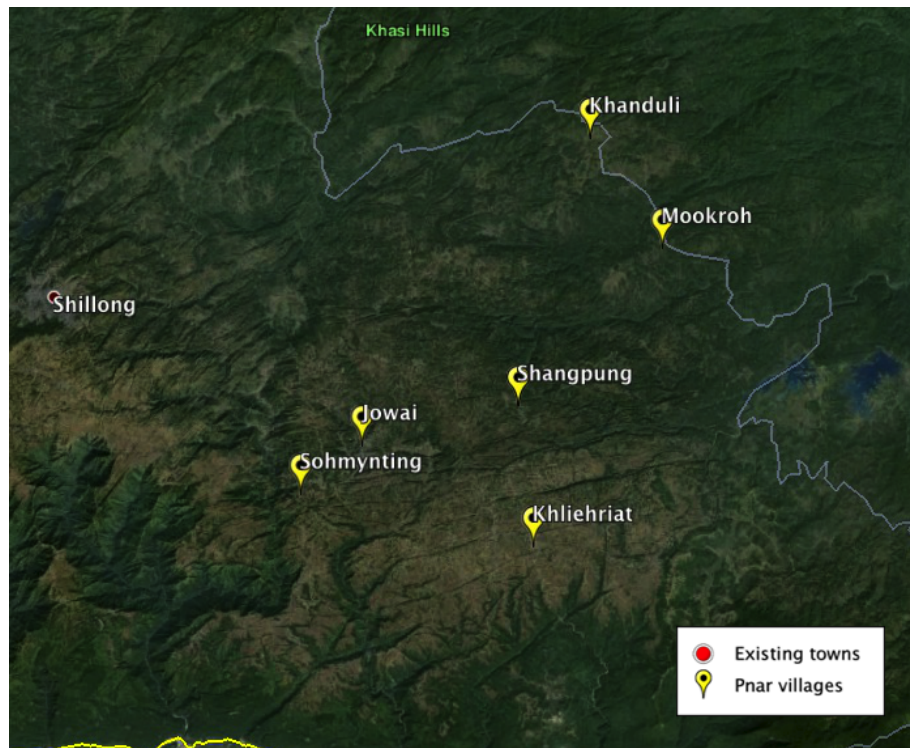


Figure 4. Some Pnar villages as points (Basemap © Google Earth, rest © Ring)

²³It can be noted that the polygons drawn here in Figures 5 and 6 have odd corners. This is partly attributable to the difficulty associated with using a mouse to draw the points. One promising way that the authors have used to get around this difficulty is to load Google Earth on a touch-sensitive device with a stylus, such as the Microsoft Surface Pro. Polygons and lines produced with the stylus are much closer to the lines drawn by hand on paper.



Figure 5. Pnar villages collectively represented as a polygon (Basemap © Google Earth, rest © Ring)

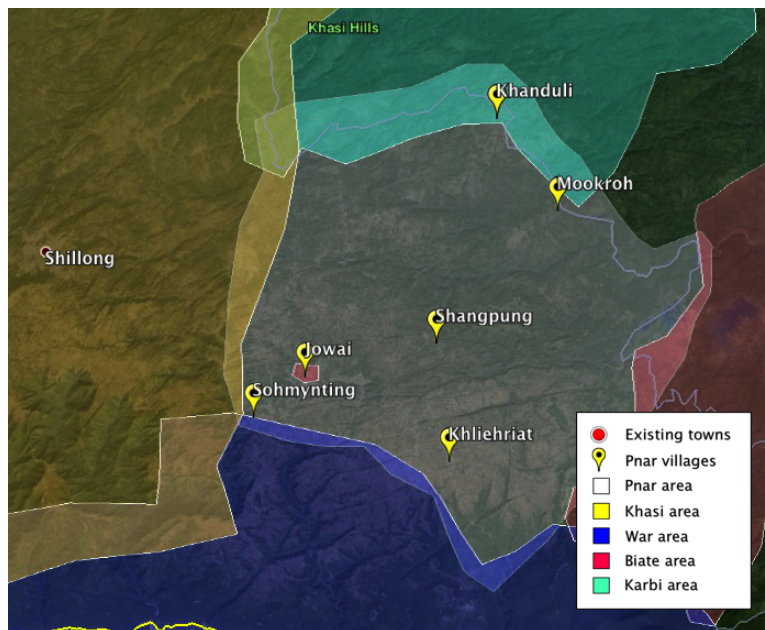


Figure 6. Some Pnar villages represented as points and a polygon in contrast to polygons of neighboring languages (Basemap © Google Earth, rest © Ring)

2.5 The challenges of borders in language mapping There are challenges with assigning languages to space and delimiting between languages (Williams & Ambrose 1988). One particular challenge is how to represent the borders of different language groups. Unlike natural geological features, which are inherent to the landscape, and geopolitical borders, which are measured with great precision, the fluidity of language and the nature of language contact can make representing the edge of a language group a particular challenge.

Mapping methods with one-language-per-area and clear borders between language groups are at odds with our understanding of the way languages can often co-exist in an environment (Luebbering 2013:41). Dahl & Veselinova (2005) suggest that with modern mapping technology and small language speaking populations the best way to represent endangered language communities is the map at the level of individual settlements where feasible. This allows us to see the specific location that languages are linked to, rather than forcing ourselves to carve specific dominions for each language. In Figure 6, this is represented by overlapping borders, with a particularly large population of an exogamous language group being identified by a polygon overlay within the Pnar area.

Even this kind of boundary-marking may distract from the reality of language contact, bilingualism, trade, or marriage movements in an area (Mackey 1988:26; Williams & Ambrose 1988:110; Stanford 2009; Luebbering 2013:44). A focus on language or dialect boundaries distracts from the fact that there are many interesting features of the contact between language groups. Conversely, linguistic boundaries may correlate with other kinds of boundaries, which may be part of a larger story of migration, history, and trade. As a classic example from language variation in North America, Labov (2001:208) discusses the ‘Inland North’ speech area, which also serves as a boundary for a variety of other activities including migration, politics, and woodwork joining styles in house-building. For those interested in establishing isoglosses for variation boundaries within a group, Labov et al. (2006:41-44) provide a detailed methodology in the introduction to *The Atlas of North American English*.

Dynamic digital maps allow for the viewer to toggle on and off certain information, depending on how the map is built, which may allow for better representation of these factors than a static map. Building your own map allows you to represent the areal reality in the best way possible; however, representing borders will always be a particular challenge to keep in mind.

2.6 Possibilities of GIS In §3 we demonstrate how to make basic maps for locating the distribution of languages in space. This is not the only use of mapping that linguists may find useful. Geographic Information Systems (GIS) allow not only for the collection and representation of geographic data, but analysis as well. While exploration of the possibilities of GIS for language documentation projects is beyond the scope of this paper, we hope that in giving language documentation researchers a basic introduction to language mapping we can open the possibility of further collaboration with GIS researchers.

Hoch & Hayes (2010) give a good overview of GIS/linguistics research and discuss potential analytical possibilities in the field. Montgomery & Stoeckle (2013:57-64) give an updated discussion with particular reference to Perceptual Dialectology research (see p. 75 for a description of GIS visualization of aggregated PD data; see also Preston 2010 and Goebel 2010 on perceptual dialectology). Kretzschmar (1996; 2013a; 2014) provides a good overview of the analytic possibilities when working with variation data, such as the likelihood of eliciting a particular item in a particular area, and Wattel & van Reenen (2010) give an introduction to probabilistic maps in variation.

Linguists may find that working with GIS practitioners gives them new perspectives on language data in relation to other geographical features. Manni (2010) gives an example of mapping both linguistic and genetic variation, demonstrating that Dutch dialects and the distribution of genetic features show some parallels but linguistic features are not dependent on genetics alone. Inoue (2010) looks at the relationship between rail line distances and use of standardized linguistic forms in Japan, finding cross-dependency between language phenomena and infrastructure access, rather than just absolute distance between varieties.

If linguists have not made full use of the possibilities of geographic research, it is comforting to know that many geographers have long neglected the complexities of mapping language (Williams 1988:1). This situation has changed, with the development of the field of ‘geolinguistics’ or ‘linguistic geography’ (Mackey 1988; Williams 1988; Luebbering 2013; Labov & Preston 2013); however, there is still much possibility for developing geolinguistics within the domain of endangered languages. Hildebrandt & Hu (2013) offer one example of such a collaboration, creating a multimedia mapping of languages and their varieties in the state of Manang in Nepal, which is helping to visualize and re-think the complex settlement patterns and histories of Manang, as well as providing new perspectives on the vitality of the languages spoken in the area. Linguists who understand the basic process of language mapping will be better equipped for possible collaboration with GIS researchers to identify existing research or develop new research directions (cf. Chamberlain 2015, in which language shift was related to environmental watersheds using maps).

3. Making maps Above we briefly discussed how geographic maps provide a useful visual representation for language documentation. In the following sections we work through the process of making maps, from GPS data gathering and processing to using that data to build a map. Currently available tools mean that making effective maps, both static and dynamic, is a skill easily within reach of language documentation practitioners.

It is possible to make many different maps using the same dataset, so it is important to have decided exactly what you want your map to represent before starting the process. We begin with an explanation of how to gather data using GPS (§3.1), and how to make this data ready for mapmaking (§3.2). We then introduce popular software for map creation (§3.3), which vary in terms of the degree of flexibility they offer, and the amount of coding required to manipulate the final product. We

introduce a simple workflow (§3.4) using CartoDB, which allows you to add and manipulate data, and has many predefined map features. Following this we introduce a more complex workflow (§3.5), involving editing data in Google Maps and styling the final map in TileMill. This option requires use of CartoCSS, a simple markup language based on CSS and built specifically for mapmaking. We conclude with a discussion of the distribution of electronic and static maps, attribution of copyright and archiving (§3.6).

3.1 Gathering GPS data GPS data is the primary means of locating positions on Earth in the 21st century. Satellites orbiting the Earth have allowed us to identify particular points in relation to a global spatial map since at least the 1980s (see Parkinson 1996; Guier & Weiffenbach 1997; Pellerin 2006). In the last 10 years this has become more commonplace for consumers, with dedicated GPS devices becoming accessible for hikers, trekkers, and ocean and land navigation.

As noted above, our focus in this paper is to describe maps and mapmaking for linguists doing language documentation who want a visual representation of geographical spaces and would like to use open-source software or tools.²⁴ Collecting GPS location data is another facet of language documentation workflow that needs to be considered when developing and undertaking fieldwork. As van Uytvanck et al. (2008) note, it is important to consider the method of collecting this data and storing it.

In the last 5 years GPS has become an essential part of smartphone technology and is used by various apps to identify users in relation to nearby businesses and advertising, or for navigation in apps such as Google Maps, with GPS functionality generally being referred to as ‘Location Services’.²⁵ Smartphones provide a useful tool for the mapmaking field linguist. Smartphone use is increasing in developing countries (where many of the world’s unwritten and under-described languages are spoken), and carrying a smartphone will no longer make you look out of place. There are also many free GPS tracking apps that can store and display real-time GPS information. Thus, collecting GPS data that can be used in mapping does not have to be an additional task in the linguist’s workflow, nor interrupt other activities. Searching for ‘GPS’ in your phone’s app store brings up thousands of free and paid apps that can be downloaded and used for a range of purposes.

²⁴Excellent proprietary mapping tools do exist and are widely used, such as ArcGIS (<http://www.arcgis.com>), but since many of the features that linguist mapmakers will use are shared with open-source tools, we describe the latter here.

²⁵Location services are available on any modern mobile smart phone. On an iPhone, this is accessed by selecting ‘Settings > Privacy > Location Services’ and turning Location Services to ‘On.’ Then location services for individual applications can be enabled. To have photos that you take on your device automatically store GPS coordinates, for example, select ‘Camera’ under Location Services and turn on the location services for the built-in camera. Other phones may vary in how the location services are enabled for photos, but the process is similar—searching ‘How to enable location services for camera on [Name of Your Device]’ using Google should give step-by-step instructions for your particular smart-phone.

With location services enabled for photographs taken on a smartphone device, the GPS information is automatically recorded when the picture is taken.²⁶ It is often culturally acceptable and even expected in many places that a foreigner would take pictures at every opportunity, allowing for the collection of GPS locations while also taking images, although sensitivity should of course be exercised. The photographs may be used by the linguist as visual aids in reconstructing their travels for map-making, or they may form part of the final mapping output by linking the mapped points to the images in a multimedia map. This is particularly useful in ethnobotanical or local history projects. As we discussed in §2.3 the researcher must always be attentive to their ethical obligation to the community when collecting photographic as well as GPS data.

Keep in mind that this location information may be displayed or recoverable online publicly if you upload the photo to a public site such as Flickr, Facebook, Instagram, or a personal website. Here your particular privacy settings will determine the photo's accessibility. You may have to remove the GPS information from the photo if you do not want the GPS coordinates to be included, as they are embedded in the image file metadata (with 'location services' on). While many people do not mind this, it is worth being aware of, particularly if you or community members have privacy concerns.

This location information is often viewable on your device by selecting the photo and looking at its 'information' or 'details.' However, this is not always the case. On iPhone, for example, the information is only viewable on third-party apps such as 'Koredoko,'²⁷ a screen shot of which is shown in Figure 7. If you have privacy concerns, it may be good to use a similar app to check whether GPS data is embedded in your photos.

A dedicated GPS device can also be extremely useful. The benefits of a dedicated device are that they generally have longer battery lives, are more robust, and some can manipulate data points on the go. However, they are not as commonplace and as such may attract more attention in a fieldwork context than comparable use of a smart-phone. They also may require additional software to transfer the data to your computer when mapping. For those who are doing a lot of walking or who want to plot boundaries or roads manually, however, a dedicated device may be the better investment. They can also be used to demonstrate to people the task of collecting data for mapping, therefore starting the community discussion of what should be mapped. GPS trackers may allow you to mark specific locations, but may also simply auto-generate a numeric value for each of the recorded GPS points, so ensure you test your GPS device, including extracting data from it, before taking it to the field. You may need to record your own metadata for the names of the enumerated points so

²⁶At the time of writing this paper, all current iPhone models are able to collect GPS data with each image, as are mid-to-high price point Android models. Small consumer camera models like the waterproof Nikon Coolpix and Canon PowerShot and professional (DSLR) models like some of the Sony Alpha and Canon EOS series also have this feature. Video cameras with GPS functionality include the Sony Handycam models and the ContourGPS camcorder. We only expect this feature will become more ubiquitous, but it is worth confirming that the model you want has it along with the other features necessary for your particular use case.

²⁷<http://labs.kawabatafarm.jp/Koredoko/>.

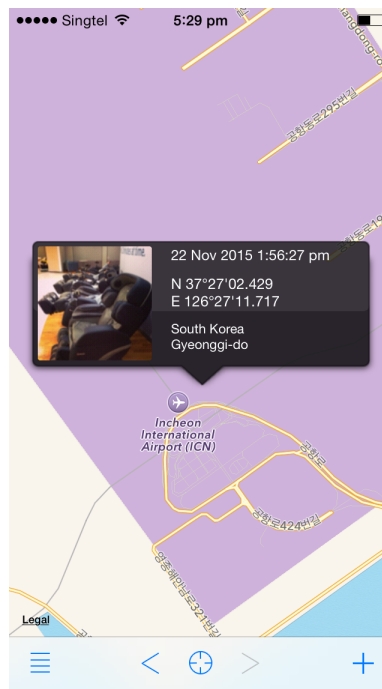


Figure 7. Screen capture of a photo with its GPS point, viewed in the Koredoko app

that you can remember what they refer to when you return to your mapping data at a later date.

It should be noted that GPS data exists primarily as points or lines. Keeping this in mind can help the linguist decide what kind of GPS data (and how much of it, i.e. how frequently points should be taken) is necessary to help identify the points and boundaries for creating a map to display the kind of information that the linguist is interested in. For example, creating a map of villages where each village is represented as a single point will require only one GPS data point per village, while collecting data on routes between villages will require a lot more GPS data points to be taken in order to produce an accurate map.

It is also possible to take more of an armchair approach to aggregating GPS data. Services like Google Earth (discussed below) now offer high-quality satellite images of many places on the earth's surface. Remote hillsides may not have the same detail as a major urban centre, but once basic GPS locations have been established it may be possible to mark out other villages, landmarks or features based on the satellite photography and familiarity with local landmarks rather than using a GPS device in situ. New technology also allows scans of (older or hand-drawn) maps to be aligned with existing coordinates,²⁸ which could become a base map for future work.

²⁸See: <http://www.davidrumsey.com/> (Georeferencer App) and the 'georeferencer' plugin for QGIS, i.e. http://www.qgistutorials.com/en/docs/georeferencing_basics.html
Also: <http://spatial.scholarslab.org/creating-gis-datasets-from-historic-maps/>. This tutorial describes working with ArcGIS, which is not free or open-source software.

3.2 Uploading, editing, cleaning GPS data The process of obtaining GPS data from a device will vary from product to product, but many will have support software, or online documentation for how to do this. Whatever the file type the GPS program works with, you should find a workflow where the final product is a KML (or sometimes a KMZ) file. KML is an acronym for ‘keyhole markup language,’ a specific arrangement of XML data that many of the mapping programs use to store GPS coordinates and other information. If you open a KML file in a text editor you will see XML tags being used to organize the data. A KMZ file is a zipped file that contains a KML and additional supporting files. This zipped file is often smaller than a KML because of the compression, and many programs that can read KML also support KMZ.

When managing mapping data you should think in terms of layers. ‘Layers’ refer to demarcations that you intend to build into your map. These are collections of similar information that you will color or shade the same, in order to contrast with other layers. Layers are the primary way that data is organized and identified in most mapping software. Maps may have a single layer that identifies oceans, another that identifies land masses, and others that identify geographical features such as rivers and mountains, or those that identify political boundaries such as country, state, and local borders. Geographical layers may be shaded or colored to give the impression of elevation, while political boundaries are often lines of different width or type.

Thinking in terms of layers will help you identify and separate the kind of GPS information you want to represent on your map. As mentioned in §3.1, this will likely be a series of points to begin with. However, you may need several different sets of points—a set of village points for one language variety, a set of village points for another variety, a line or periodic points to identify potential trade routes, or a set of points to identify marketplaces or cultural/historical points of interest. With careful metadata assisted by photos that you take along the way, these GPS coordinates become a rich source of information to support future maps and mapping projects. If you take points without photographs, remember to make note of the location of each point (for example the village name or the route name).

An alternative to using KML files is to create a list of longitude and latitude values for points you want to map and save them as a CSV (comma separated values) file in a program like Excel. Each layer of information you want to display should have its own CSV file. You may already have these values, or have taken them from Google Maps or another mapping platform. An example for locations in a CSV is given in Table 1, which lists the location of the four main Yolmo language areas in Nepal. The X column is for longitude and the Y column is for latitude. These points were used to generate the map in Figure 3 and are also in the downloadable CSV file ‘Yolmosites’ that accompanies this paper.

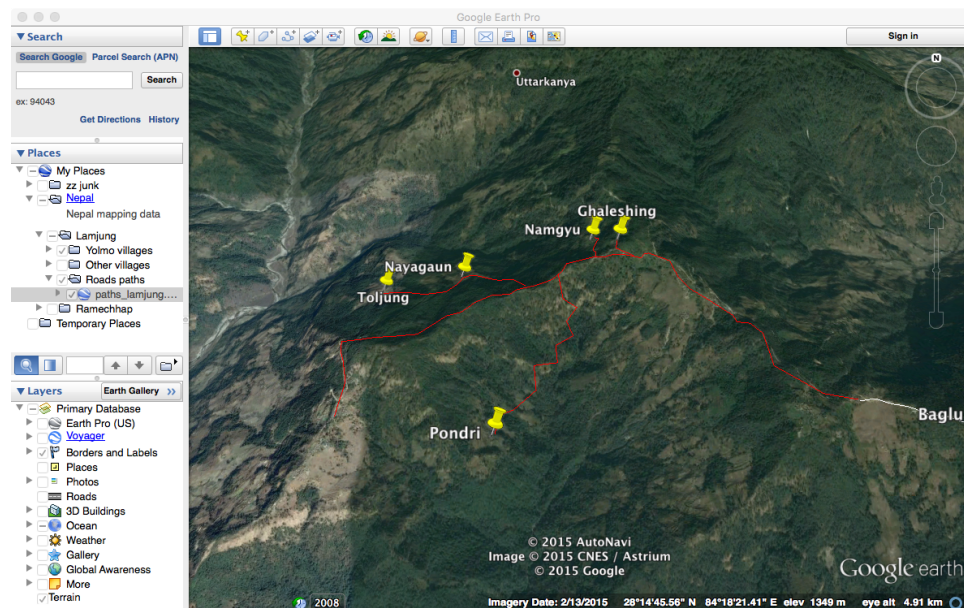
CSV tables can only be created for point data, not for line or polygon data. If you have a larger number of points it may be more time-consuming to generate a CSV table than working with GPS and a program like Google Earth.

Table 1. Set of longitude and latitude values for locations where Yolmo varieties are spoken

Language	X	Y
Lamjung Yolmo	84.316263	28.234986
Melamchi and Helambu Valley Yolmo	85.589647	27.949222
Kagate	86.071987	27.349813
Ilam Yolmo	88.040085	26.933091

One of the most useful tools for uploading, editing, and cleaning GPS data is Google Earth, a free cross-platform program.²⁹ Google Earth will accept data from a wide range of formats, which can then be manipulated, and allows KML export. Many GPS devices will allow direct upload to Google Earth. Google Earth allows you to move data points, create data points and export these for use in other programs. Other programs can also be created as add-ons to Google Earth.³⁰ It is possible to take a screen capture of Google Earth to use as a display map; however, it is not a clear or elegant way to represent geodata and therefore we only use it as a first step for editing and manipulating.

Figure 8 is a screenshot of Google Earth Pro, which is now freely available for consumers, with Lamjung Yolmo village data visible. Google Earth Pro allows for higher resolution screen shots, and the ability to create videos of movement through the mapped environment.

**Figure 8.** The Google Earth representation of Yolmo village GPS data in Figure 2

²⁹<http://www.google.com/earth/>.

³⁰Such as those found at: <http://earth.tryse.net/programs>.

The panel on the left includes lists of GPS data, which can be arranged thematically. You can observe that the Lamjung Yolmo village points are kept together in one folder (“Yolmo villages”), alongside other villages and roads traced either with a GPS unit or by looking at high-quality Google Earth satellite imagery. The Lamjung folder is next to a Ramechhap folder, which contains locations of the Kagate-speaking villages in another district.

Figure 9 shows the main features of the Google Earth toolbar for creating map data. The red rectangle highlights elements that can be added to the current view. From left to right: points, polygons, paths, and image overlays, which each create a new layer. The final image of the camcorder allows for the creation of ‘flyover’ tours of areas, which may be useful for some illustrations of linguistic areas.



Figure 9. Screen capture of a Google Earth toolbar

Google Earth allows you to move data points, create data points, and export these for use in other programs. It is possible to take a screen capture in Google Earth. Using the small ‘printer’ icon on the top bar, the program will allow you to create a title and a legend.³¹ The screen is then saved as a PDF. The Google Earth logo will be created on the map and cannot be removed.

Figure 10 is a screenshot of the Google Earth representation of the Yolmo villages in Figure 2 that could be exported for publication. You can see clearly where Google Earth has stitched together different GPS satellite images.

One advantage of displaying maps in Google Earth is the impressive representation of the three-dimensional environment. Figure 11 is the same map data, rotated to demonstrate more clearly the distribution of these five villages on the hillside, and the nature of Lamjung’s mountainous terrain.

If you have taken pictures with embedded GPS data, you can download or transfer your pictures to your computer, ensure that they contain GPS data, and use a program such as Geotag³² or Photo KML³³ (both free), to convert your images into sets of GPS points that you can open in Google Earth or another map editor. Geotag is a Java-based program for getting GPS data (in EXIF format) from images and exporting the GPS information into several formats, including KML. To use it, simply download the Java program from the SourceForge site (link in footnotes), make sure you have the latest version of Java, and run it. To add data select ‘File > Add image...’ or ‘File > Add images from directory...’ (Figure 12). To export data, select the files that you wish to group together, right click, and select an export option such as ‘Google Earth

³¹This process does not always work as desired, which is why the legends in Figures 4–6 were created manually.

³²<http://www.geotag.sourceforge.net/>

³³<http://www.visualtravelguide.com/Photo-kml.html>

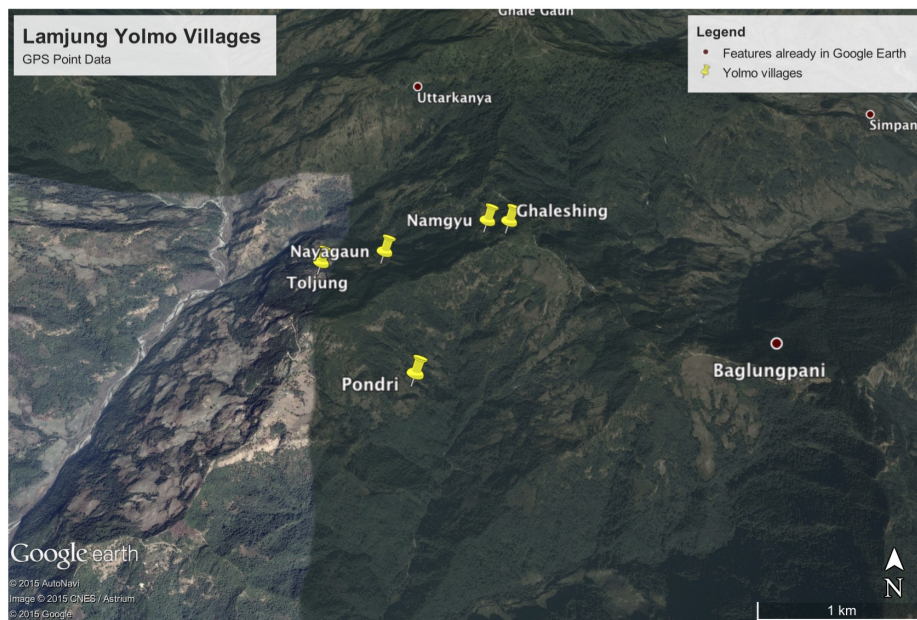


Figure 10. The Google Earth representation of Yolmo village GPS data in Figure 2

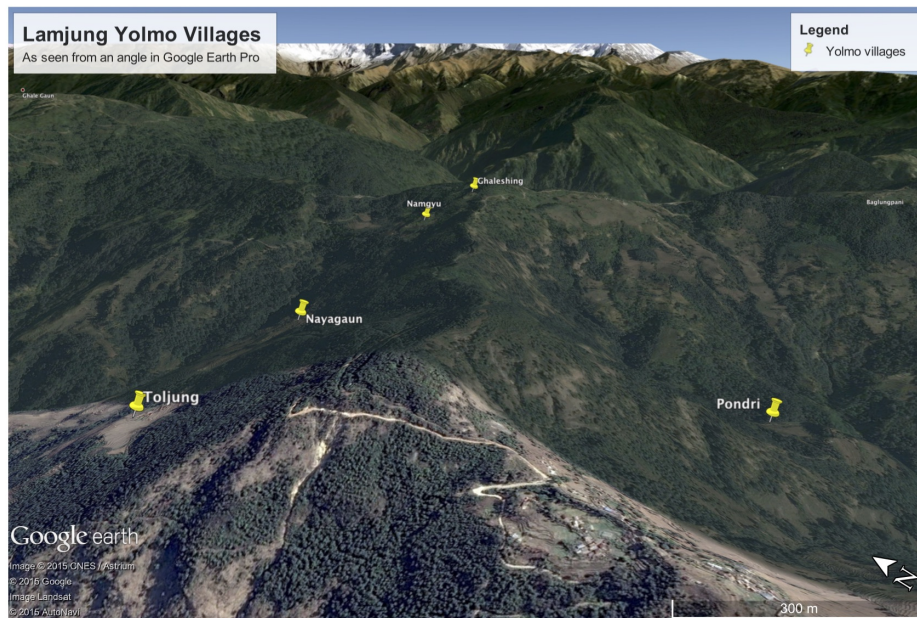


Figure 11. The Google Earth representation of Yolmo village GPS data in 3D

> Export selected images' (Figure 13). Then select a location on your computer to store the data and type a name for the KML file.

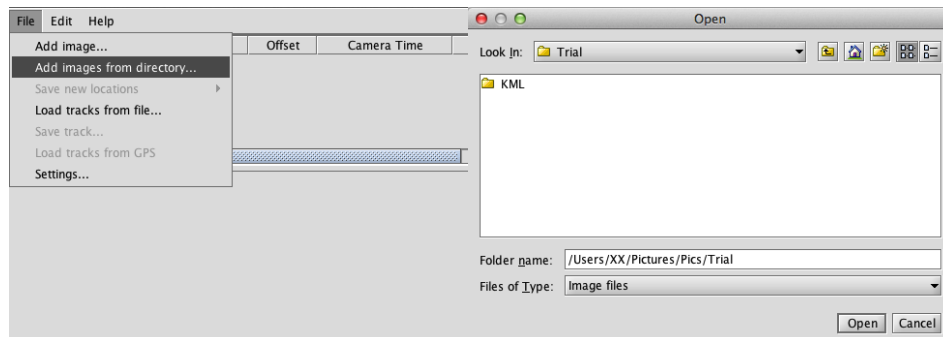


Figure 12. Opening images in Geotag

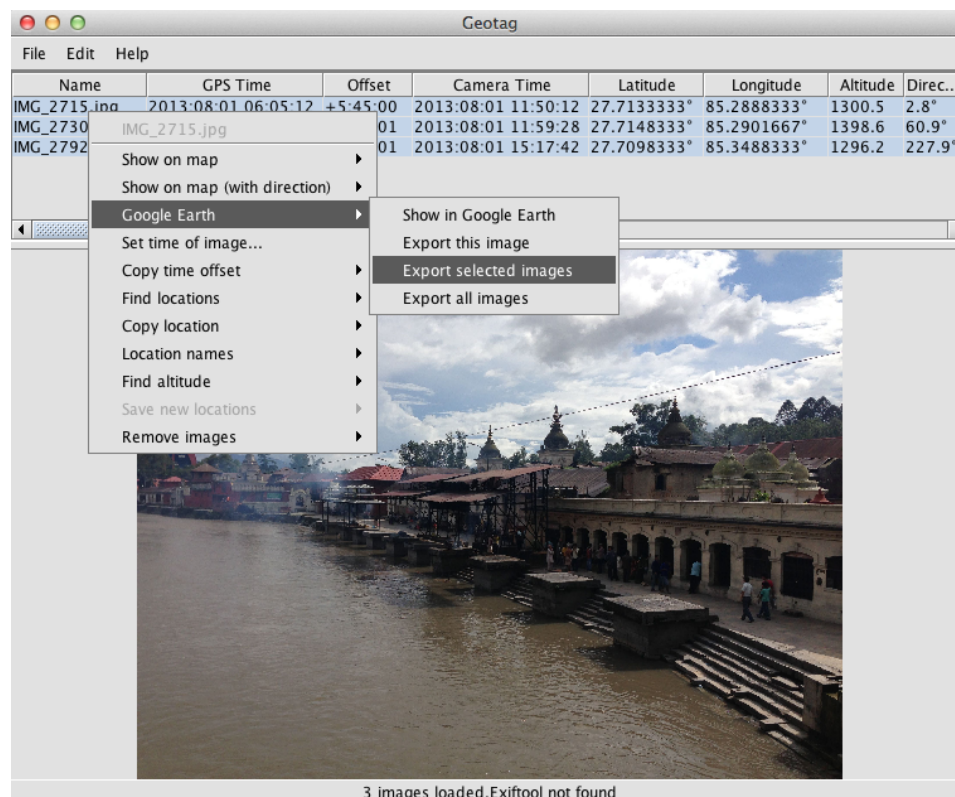


Figure 13. Exporting image information from Geotag

There are a number of steps you can take to make it easier to create geospatial datasets from GPS photodata. The first is to organize your photos into folders. Create folders for each layer that you want to build into your map, and copy or move the photos that are associated with that layer into the proper folder. Then open each

folder individually within Geotag and save each group of images in the Google Earth format. Here you will notice that the KML file links to the location on your hard drive where your image is stored, and that all the images in a layer are grouped together under the heading ‘Geotag.’ This is information that you can edit in Google Earth. You can change the names/titles/headings, delete the images, and even adjust the location of each of the points. To export from Google Earth, the easiest process is to right-click on the point (or folder of points) you wish to export, select ‘email,’ and email them to yourself (Figure 14). They will then be in an email as a downloadable KMZ attachment.

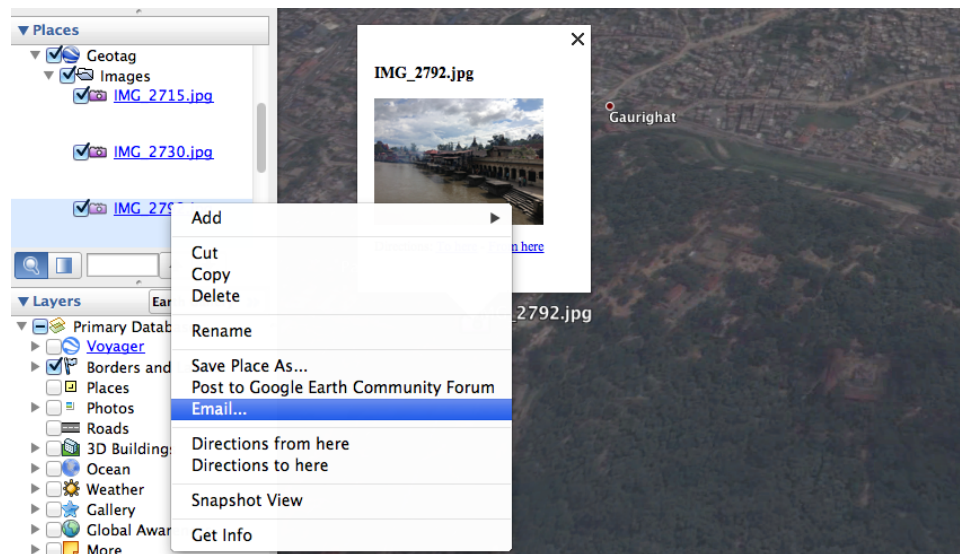


Figure 14. Exporting from Google Earth

Although there are some advantages to using Google Earth, it is not always the best way to create a map for illustration or display. As you can see, the pushpins are not very attractive,³⁴ and while the satellite images of the terrain look fine on screen, it may be too much information for other mapping purposes or for black and white publication. Therefore we mostly use it as a first step for editing and manipulating.

3.3 Mapmaking using existing software In this section we identify some of the useful mapping software that is freely available. Some of these require an internet connection, while others can be downloaded for use offline. These programs also vary in terms of user friendliness, i.e., visual simplicity, variety of menu options, format of data, intuitive nature of interface, and access to underlying code. The software that we review briefly here are (in order of complexity for the user from simplest to most technical): Google Earth, CartoDB,³⁵ TileMill,³⁶ and QGIS.³⁷ This section is intended

³⁴The pushpin style can be edited by right-clicking, but there are a limited number of options.

³⁵<http://cartodb.com>.

³⁶<http://www.mapbox.com/tilemill>.

³⁷<http://www.qgis.org>.

as a basic outline of the kinds of tools that exist, and the advantages and disadvantages of each project from the perspective of linguists interested in mapping. It is not intended to provide a comprehensive review. More detailed walk-through tutorials are given for CartoDB in §3.4 and Tilemill in §3.5. These are not the only available mapping programs available; amongst many others, Story Maps³⁸ allows for the creation of map-based infographics, and Mapbox³⁹ is the fully online successor of TileMill.

Google Earth, already introduced in §3.2 above, is freely downloadable software for visualizing the earth as a globe and for zooming in to specific points or geographical features. It incorporates GIS data and allows users to add and share points and descriptions, and is particularly useful for identifying places you have been on the globe through geographical features. It has some of the most detailed satellite imagery available to general users. It is possible to import, draw, and export points, lines, and polygons in KML and KMZ formats. For those who have an internet connection but do not want to install Google Earth, Google also offers My Maps as part of Google Maps.⁴⁰ This has fewer features than Google Earth but allows for the creation of points or lines in layers that can then be exported to KML format. Google Earth and My Maps also allow you to share your maps publicly within these programs. TileMill and CartoDB are the two programs we have had the most experience with, both in our own research and in training other mapmakers. They build on the same fundamental skills as other mapping programs, but represent two ends of the scale in terms of ease of accessibility.

CartoDB is an online map-making tool, primarily for dynamic digital maps, although they do also have a static map export option. Because the maps are hosted on the CartoDB website they can be accessed anywhere and embedded into other websites. CartoDB requires users to register an account, with the free user accounts limited in their storage size. Maps made in the free accounts are currently limited to only four layers, which will constrain the number of features you can represent on a single map. All maps made with the free version are publicly viewable, so may not be appropriate for mapping sensitive data. Paid accounts offer more storage and private maps. CartoDB allows you to upload existing datasets, for example those created in Google Earth, but also to create or modify datasets within the program, which is not possible in map design programs like TileMill. The data created or modified in CartoDB can also be exported again. CartoDB offers design interfaces that offer less flexibility than fully featured software but allow for the quick creation of maps with no knowledge required of scripting languages. CartoDB have a version of their free accounts for researchers with an academic email address.⁴¹ This allows for live-update linking to a Google Docs form, which may be useful for data collection and update.

³⁸<http://storymaps.arcgis.com/>.

³⁹<http://www.mapbox.com/>.

⁴⁰<http://www.google.com/mymaps>.

⁴¹<http://cartodb.com/industries/non-profits/>.

TileMill is specifically designed with the non-specialist cartographer in mind, and focuses on allowing users to create visually pleasing maps. Unlike CartoDB, it is not possible to edit datapoints within the program; they must be edited in another program such as Google Earth. Built mainly to create maps used in web browsers, a coding window allows users to fine-tune adjustments to layers using a variety of CSS called CartoCSS. While it is not necessary to fully learn this code, users may find that learning the basics will make their maps more visually interesting and understandable. Tutorials are available online,⁴² and it is possible to follow increasingly difficult tutorials working with provided data to master the features. The data that you use to make maps is stored locally on your own computer, which means TileMill requires you to set up your own hosting if you want the maps to appear online.

With a focus on creating interactive web-based maps, TileMill is not always ideal for users who are primarily interested in creating print-ready maps, as exporting may require some finessing. As another caveat, TileMill is no longer in active development, with the developers having moved to development of Mapbox Studio.⁴³ TileMill is currently easier to use than MapBox Studio, offers better static map export, and we believe it is still preferable for the kind of language mapping work linguists are likely to undertake. One limitation is that TileMill does not work on Mac OSX 10.10 (Yosemite) and above.⁴⁴

QGIS is a full-featured professional graphical GIS editor.⁴⁵ This program is freely available and runs on all major operating systems. It imports from nearly any mapping format and allows the user to edit and create GIS layers of points, lines, and polygons in a visual layout similar to Photoshop. However, the level of detail and array of menus and options outstrip most basic mapmaking needs. This is an excellent tool but is much more powerful than the typical linguistic mapmaker requires. We do not include QGIS in our workflows, as we find that the less powerful programs are still more than sufficient for the needs of most language mapmakers. If you are interested in building more complex GIS projects, learning QGIS may be worth the time, but starting with the programs we discuss will offer good basic training in mapping nomenclature and design. QGIS has additional features beyond the scope of tools such as TileMill and CartoDB. One particularly useful feature is the ability to align existing printed maps with geographical coordinates in a process known as ‘georectification’.⁴⁶

There are many different ways to create maps, but as Kretzschmar (2013a:67–68) notes, “the worst kind of linguistic map is the one that the analyst cannot make, and so cannot look at or show.” In this section we describe two approaches that we have found to be both useful and accessible. The first is a ‘simple’ workflow using CartoDB

⁴²<http://www.mapbox.com/tilemill/docs/manual/>.

⁴³<http://www.mapbox.com>.

⁴⁴People have had success running TileMill on Mac through Wine (<http://www.winehq.org/>) or Docker (<http://hub.docker.com/>); however, as both of these require familiarity with Terminal commands we cannot readily recommend these options to all readers. TileMill servers can be created, where the user logs into the program through a web browser, which avoids the limitation for Mac users. It may be possible to talk to your University IT department about setting up such a service.

⁴⁵<http://www.qgis.org>.

⁴⁶A tutorial on how to do this can be seen here: <http://www.youtube.com/watch?v=A-jBYc9pLiQ>.

that allows the user to create a basic map easily and also enables further expansion and collaboration, but which requires an internet connection. The second is a more complex workflow that makes use of Google Earth and TileMill.

3.4 Basic workflow: CartoDB CartoDB allows users to either import existing data in a range of formats, including KML and CSV, or to create and edit datasets within the program. In this section we will walk through creating a basic map of Yolmo language locations using the CSV sheet given in Table 1. Addressing Girth's (2010) four questions in §2.4 the aim of this map is to represent the location of the four known Yolmo varieties (q. 1). These will be mapped, as will the geopolitical border of Nepal and the relevant major cities (q. 2). The border will be mapped using line data, while the language varieties and cities will be mapped using points (q. 3). The mental picture that this map aims to create is an understanding of the location of these varieties in relation to each other, the country of Nepal and major cities (q. 4). That dataset can be downloaded alongside this paper, and is called 'Yolmosites'.⁴⁷ Locations for Kathmandu and Pokhara are also mapped, as they are key locations in Nepal; these are available for download as 'Cities' from the same location.

You will need to create a user account before you begin, and all of your data will be stored online, so if you create or modify data in CartoDB, remember to download the data occasionally for backup and archiving. CartoDB allows you to create datasets, with each representing one layer of information on a map, and then to combine multiple datasets into maps. Figure 15 is a screenshot of the opening page when you log into your CartoDB account (the maps are from Gawne's account).

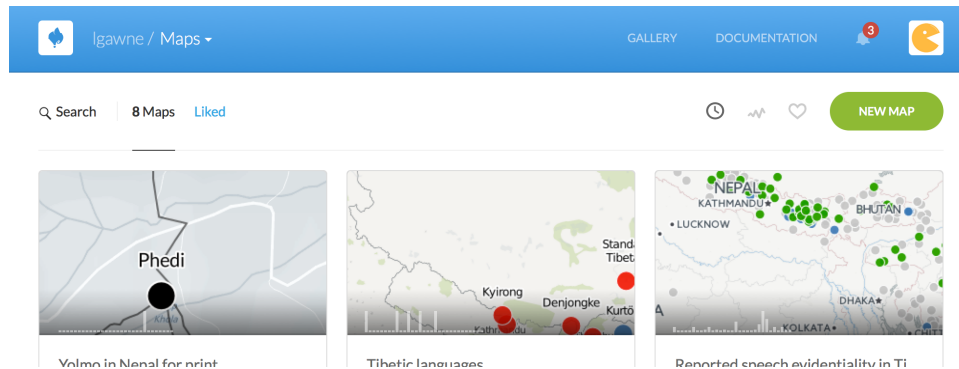


Figure 15. CartoDB list of maps in Gawne's account

Clicking on 'maps' at the top left will allow users to toggle over to the page where their map data is stored. Datasets provide the basis of creating different maps—multiple maps can draw on the same datasets, allowing you to create different visualizations of the same data very quickly. It is possible to view each dataset as a separate map as well as building them into combined maps.

⁴⁷The datasets are available for download from
<https://scholarspace.manoa.hawaii.edu/bitstream/10125/24692/2/Yolmosites.csv> and
<https://scholarspace.manoa.hawaii.edu/bitstream/10125/24692/3/Cities.csv>.

When you create a new map (green button at the top-right), CartoDB gives you the option of creating a blank map, or using one or more of your datasets. Datasets can be imported in multiple formats (including KML and CSV) or you can browse the datasets available on the website, although their datasets may not match your needs. Gawne had already uploaded the Yolmo sites dataset to her CartoDB account, so she selected the ‘yolmosites’ dataset. She also selected a dataset called ‘cities,’ which is a CSV with the locations of Kathmandu and Pokhara, as these provide useful references for people familiar with Nepal. This dataset is available in the CartoDB database. When datasets are selected and the map is created, CartoDB will provide a default styling to the points on the newly created map. Figure 16 is a screenshot when the map has been created.⁴⁸

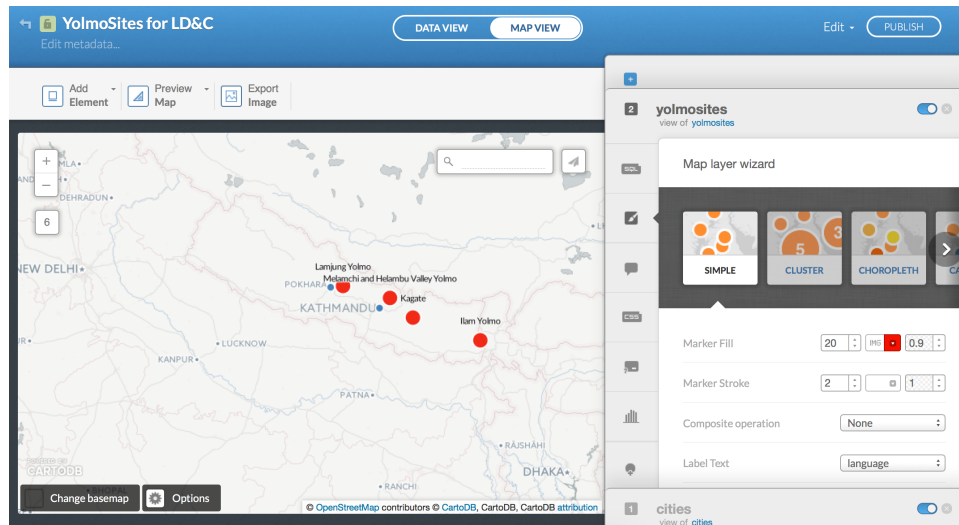


Figure 16. New map with Yolmo sites and cities datasets selected

Moving around the screen from the top in a clockwise direction, we see that ‘Map view’ is currently highlighted. It is possible to also select ‘Data view’ which will show a table of values for whichever layer you currently have selected. Each dataset layer is represented in its own tab on the right of the map visualization. In this screenshot ‘yolmosites’ is currently open, and ‘cities’ is stacked at the bottom. The toggle at the top of the tab can hide that layer from the visualization. Currently displayed in the ‘yolmosites’ tab is the layer wizard, which is used for basic styling of the dataset. Other useful features along the left side of the tab include a CSS window for more detailed styling of data using CartoCSS code, which we discuss further in the workflow below (§3.5). It is also possible to enter SQL code or filter the data providing basic analytical tools, as well as creating legends and info windows for digital map display. The options tab allows you to add features for digital display of maps, including a share function, and locking the zoom if you only want people

⁴⁸The completed map can be viewed at http://lgawne.cartodb.com/viz/c5d973ee-9342-11e5-bd91-0e5db1731f59/public_map.

looking at the map at a particular zoom level. Next to this is the basemap option, discussed below. ‘Publish’ in the top right is for digital distribution and allows you to send a link, or embed the map in a website, while ‘Export image’ on the left of the screen allows for a static download as either a JPEG or PNG image.

The first design choice is to select a basemap. CartoDB comes with a variety of basemaps, including simple geo-political maps, satellite data, and styled maps; you can change between them in the ‘change basemaps’ menu on the bottom left of the map. It is possible to import other backgrounds, but it is easier to select one of the options already loaded. CartoDB basemaps cannot be edited. You may find that there are labels or features that you would prefer were not shown. We suggest you attempt to find one of the available basemaps that best meets your needs, but if this is insufficient you may want to consider using TileMill, which allows for complete control over all design elements.

One thing that should be noted about CartoDB is that different basemaps are licensed from different organisations and will have different usage limitations. We recommend using basemaps from Open Street Map, like in Figure 5. The OpenStreetMaps cartography is licensed under the Creative Commons Attribution-ShareAlike 2.0 license⁴⁹ (CC BY-SA). This means you can copy, distribute, transmit, and adapt the data as long as OpenStreetMaps is given attribution. Maps provided by Nokia in CartoDB only allow for personal non-commercial distribution, which may not be appropriate for publication requirements. For this example we have selected ‘Positron (lite),’ which has minimal included labels, allowing us to choose what is shown. Zoom in until you are at a level of detail that you like. The zoom level and centering will be remembered by CartoDB whenever you return to the map, or embed it on another site.

Clicking on a point will bring up an infowindow. You can decide what is displayed in an infowindow, or if they come up, by selecting the infowindow icon for the relevant layer. In the editing mode the info window also includes a button for editing geometry. This is shown in Figure 17.

The selected point can then be moved around. For a line or a polygon the features that can be moved around will be rendered with similar blue points. Remember that when you edit a data point this will edit the point in the underlying dataset, modifying that value in all maps that use that dataset, so proceed with caution.

New points can be added to a layer in the map view by selecting the ‘add feature’ button at the bottom of the list of icons for each tab. If the dataset contains points, then the program will only allow you to add points; similarly, lines and polygons can only be added to line or polygon data. You may choose to create all of your data this way in CartoDB, although CartoDB maps do not have the same density of references or clear satellite images that Google Earth has.

Once you are happy with the zoom level, basemap and data that you are working with, you can use the map layer wizard to quickly style the data. The size, color, and outline of the data in each layer can be modified. Labels can also be added by selecting the value with the label you want to add. In this example the column head is titled

⁴⁹<http://creativecommons.org/licenses/by-sa/2.0/>.

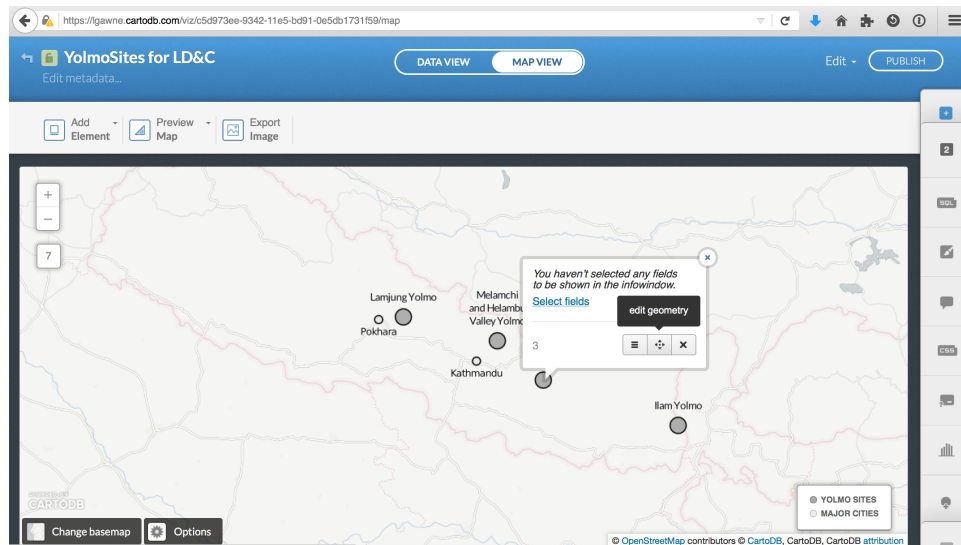


Figure 17. New map with Yolmo sites and cities datasets selected

‘language.’ The font color and style can also be modified using the variables given. More detailed styling requires the use of CartoCSS in the CSS tab. We introduce CartoCSS in §3.5. If you open the CSS tab after styling your map you will see the work you have done rendered as CSS. If you make any modifications to the CSS this will take precedence and you will not be able to use the layer wizard to further style your data. We went with a neutral color scheme to suit the minimalist map.

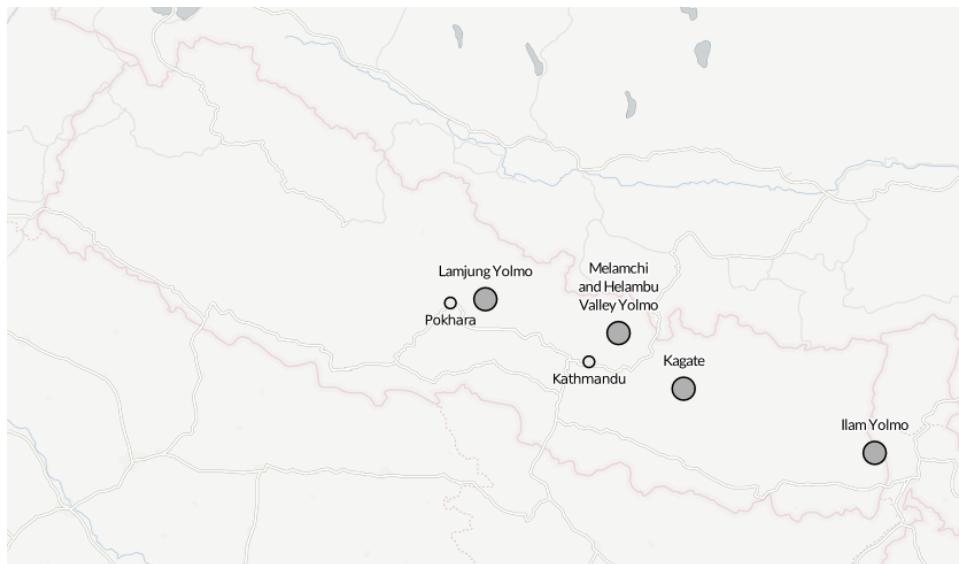


Figure 18. Map made using the CSV in Table 1 (basemap © OpenStreetMap contributors, basemap design © CartoDB, rest © Gawne)

You can export the map for the creation of a static product by selecting ‘Export image’ near the top left of the map. Figure 19 shows the export screen options.

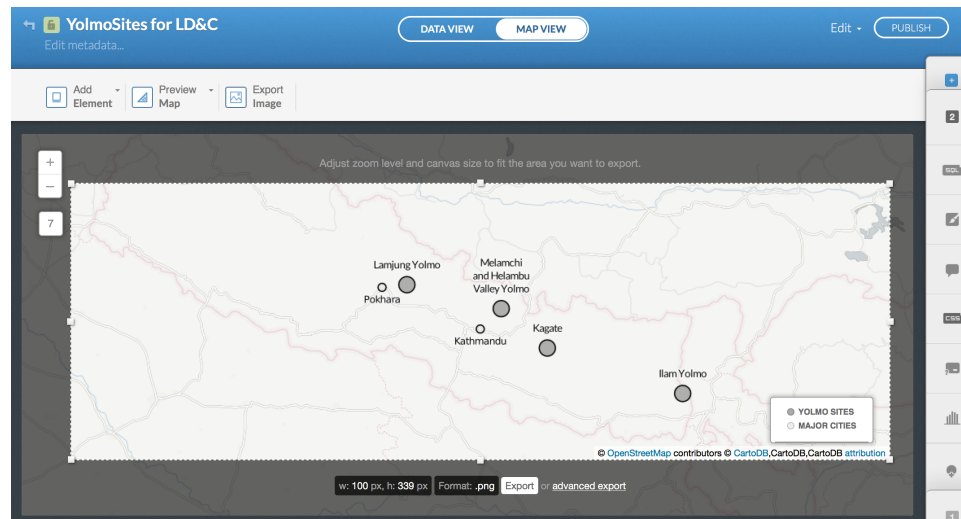


Figure 19. The export image function in CartoDB

A selection window is given, which provides a quick way to create a map if all of the relevant information fits onto the screen at the zoom level you want. Exports are either PNG or JPEG. If you need to capture a larger map than is presented on the screen, select the ‘Advanced export’ option. You can select the pixel width and height, with the export then calculated from the current zoom level and centering of the map. This export does not have the flexibility of TileMill export, and may require you to zoom in and then choose a large pixel area to ensure print quality.

If the map is to be displayed digitally, the user can also add ‘infowindows,’ which are a small amount of text that can be read when a datapoint is selected. Different options are available for hovering over a data point or clicking on it. This feature may be useful if you have plans to use a map online, but will not be available in print. This text would be an additional column in your CSV or KLM file and can be added in the data view—it may be a small description of a village, a lexical item you want to compare across different populations, or whatever you like. For this map we have created an infowindow that gives the latitude and longitude when the point is hovered over, and the latitude, longitude, and variety name when the point is clicked (Figure 20).

You can add a text legend to explain different colors used on the map by selecting ‘legends’ on one of the relevant tabs. This will only appear on the online map, not on the static export. You can see the legend we created for the Yolmo map in the above figures. Taking a screenshot of the legend and then pasting it into the exported image provides an efficient solution.

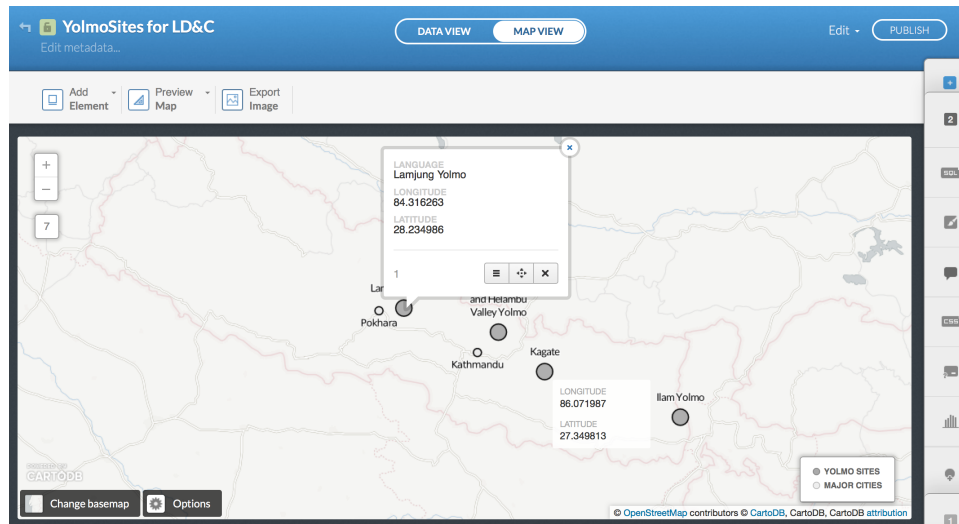


Figure 20. Infowindows in CartoDB (click on left, hover on right)

The CartoCSS coding needed to make dynamic maps is given in the later half of §3.5.⁵⁰ The basemap is a satellite image, which would be too detailed for print, but works satisfactorily for a digital map. Zooming in on the Lamjung Yolmo and Ramechhap points reveals the distribution of the villages at zoom levels greater than 10. This draws on the Yolmo village data from Figure 8. This means that instead of Figures 1 and 2 being separate maps, in a digital map they become different levels of the same map in Figures 21 and 22. The map can also be enriched over time. Mapping data for Kagate villages was added more recently, and data for the other communities in Ilam, Melamchi and Helambu can be added when it is acquired at a later date.

CartoDB can make intelligent guesses about different kinds of maps you may wish to make and can use the wizard features to style such maps. This includes density heatmaps, category maps, and choropleth maps, where areas are shaded or patterned in proportion to a statistical measurement. When you import a dataset and view it in map mode, CartoDB will inform you if it thinks it can create a particular map type with a popup in the bottom right corner of the map. Below are some examples of these map types using data from a research project Gawne is undertaking on reported speech evidentiality in Tibeto-Burman languages. The point data, available from Glottolog as a template base, was combined with information about evidentiality from the literature. Figure 23 is a map with the current status of information on each language.⁵¹ This is a category map, with one column of data indicating whether a reported speech evidential particle is present, not present, or if there are insufficient references to determine.

⁵⁰To demonstrate the advantage of using CSS styling, a dynamic map using the same data can be found at http://lgawne.cartodb.com/viz/94f75f72-a200-11e4-90b6-0e0c41326911/public_map.

⁵¹The most up to date version of the map is available at http://lgawne.cartodb.com/viz/95fc31aa-3543-11e5-bc55-0e0c41326911/public_map.



Figure 21. Map of Yolmo in Nepal, zoom level 7

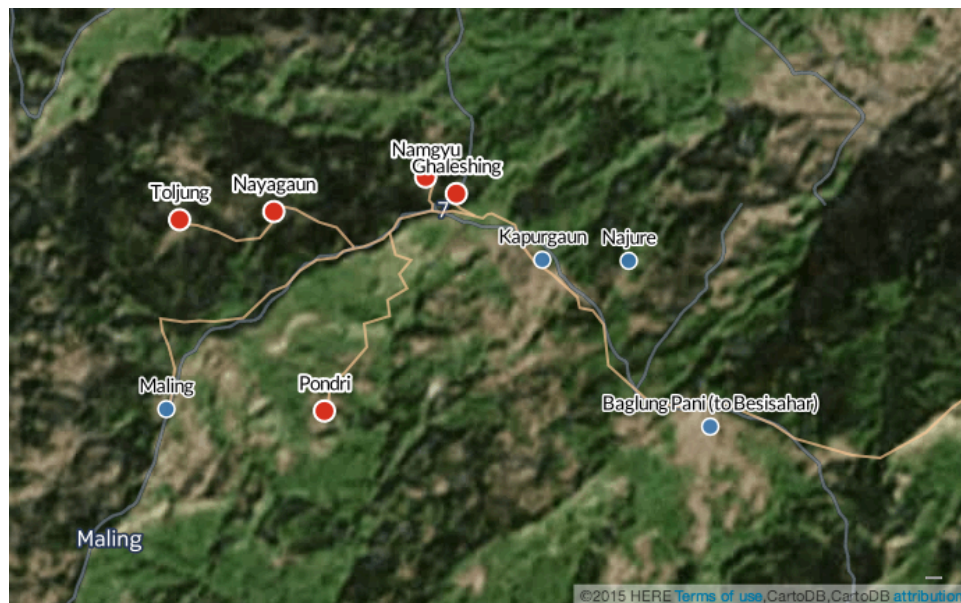


Figure 22. Map of Yolmo in Nepal, zoom level 14

The same dataset can be used to generate other data. Figure 24 is a heatmap, showing the density of points in the map. This type of map is not particularly appropriate for this kind of work, as each point represents greatly varying populations; for example, Tibet is under-represented due to low population density, and no map data is available for many under-researched Tibetic varieties. It does, however, nicely show the general area in which Tibeto-Burman languages are found.

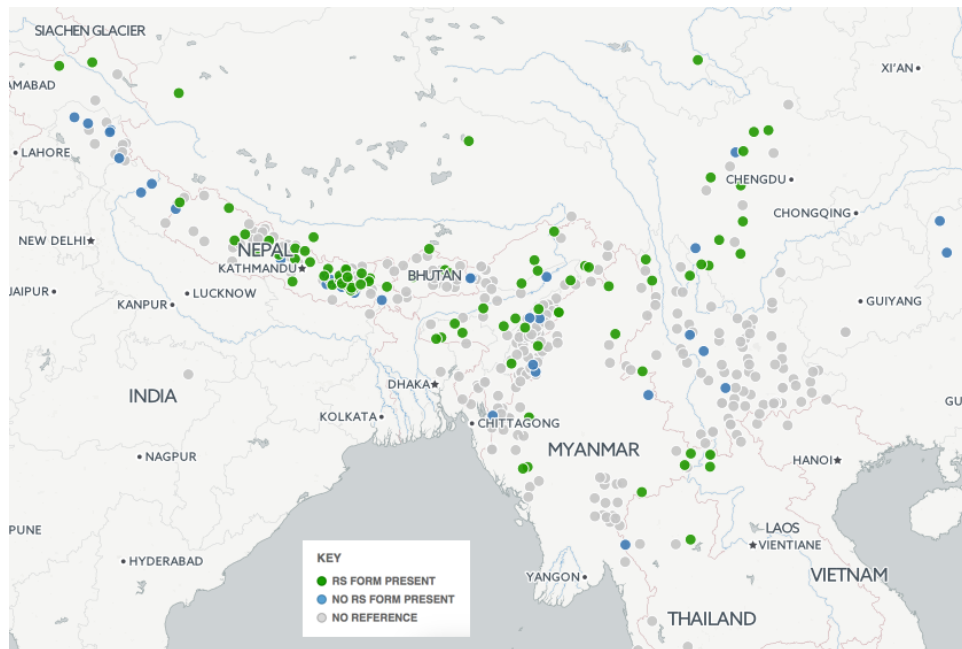


Figure 23. Presence of reported speech evidentiality in Tibeto-Burman languages

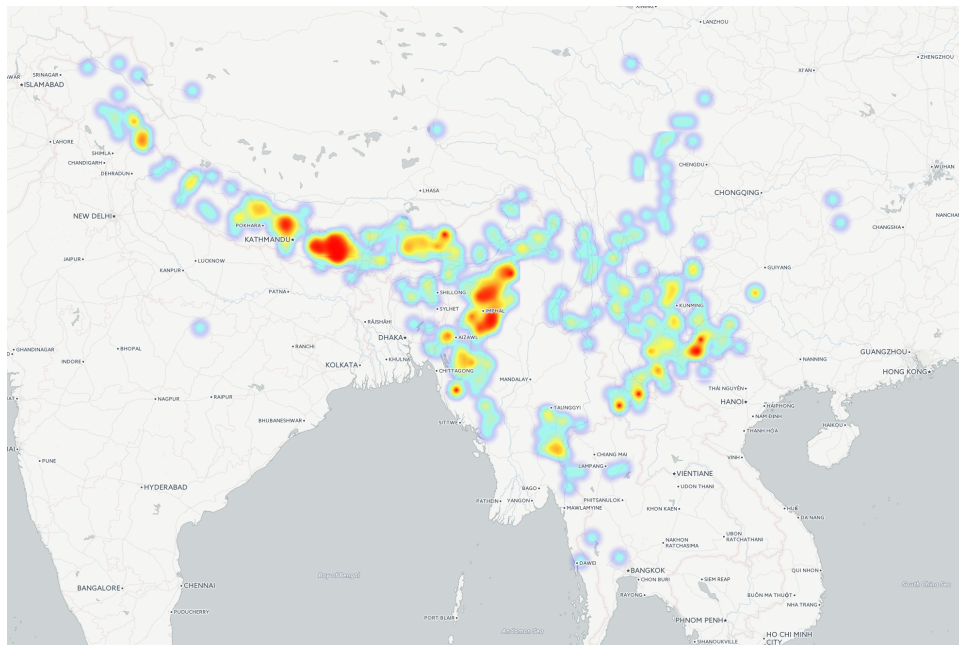


Figure 24. Heatmap of Tibeto-Burman language data

The final map we will demonstrate for the Tibeto-Burman data is a choropleth map. These maps show shading or color change in relation to a particular value. The

map in Figure 25 is slightly facetious, as it is plotting latitude values—obviously, as we move further north on the map the latitude numerical value will be higher. It does nicely demonstrate how the color patterning works. With additional data added to this set it would be possible to do a choropleth for data such as population size or language vitality scores.

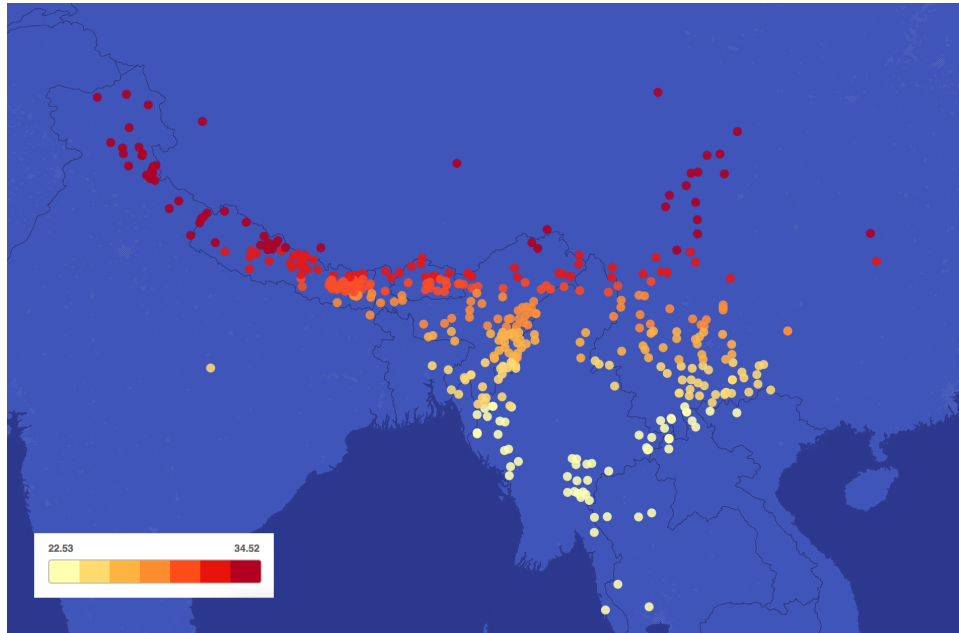


Figure 25. Choropleth map of Tibeto-Burman language data showing latitude

3.5 Complex workflow: Google Maps and TileMill This workflow describes a method of using My Maps in Google Maps or Google Earth and exporting data to TileMill. We consider this to be a more complex workflow as it requires moving between Google Earth for data editing and TileMill for map styling, while it is possible to do both in CartoDB. It is also more challenging, as TileMill does not have the same graphical interfaces for data styling as CartoDB and, instead, users must use CartoCSS to manipulate the design of the data. This workflow will draw on the ‘Yolmosites’ and ‘Cities’ CSV files again, as well as some publicly available data. The final CartoCSS code is given in Appendix 1 if you want to build and play with this map. Although TileMill is no longer in active development, there is still a large collection of useful materials on working with the program, and with CartoCSS, both on the MapBox website (creators of TileMill),⁵² and the MapBox GitHub site.⁵³ This walkthrough is not exhaustive by any means, but will provide the basic skills necessary to building a language map, as well as introducing some of the skills needed to build a dynamic map.

⁵²<http://www.mapbox.com/tilemill/docs/crashcourse/introduction/>.

⁵³<http://github.com/mapbox/cartto/blob/master/docs/latest.md>.

The first step is to get your data into Google Maps or Google Earth. This can be done simply by adding points or polygons to a map in My Maps or to a globe in Google Earth. If you have a user account with Google (which can be created easily and is also free), you have the option of saving the map so that it can be accessed from anywhere. Adding points or polygons to the Google map is a relatively simple matter of moving the map until you are viewing the area to which you want to add point or polygon data, and then using the tools in the toolbar. In §3.2 we illustrated some options for getting GPS data from your photos tagged with location data by your smartphone and importing this into Google Earth. Figure 8 is given again here as Figure 26 to illustrate.

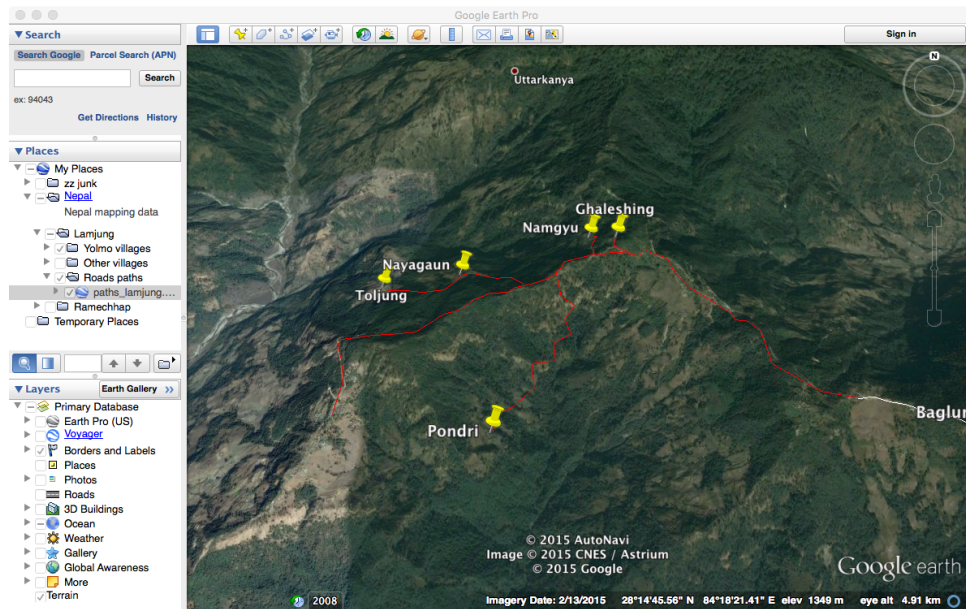


Figure 26. Google Earth interface

The yellow pushpin and ‘+’ on the top bar allow you to add a point. Next to that are options for adding a polygon, a line, and a layer, as well as making an animated ‘flythrough’ of the map. When you are happy with the points you want to map, you need to export the data as a KML. The envelope allows you to select the relevant data to export in an email, which is the easiest way to get data out of Google Earth.

Ensure you save this KML file in a stable location that you can remember. Every time TileMill opens the project it will go to this location to find the data to render. TileMill does not import the data that it uses to build maps, but reads it from the location it is saved every time the program is opened. TileMill installs with a number of projects included. These projects can provide a handy guide for various styling choices, and it is worth having a look through them. Tilemill organizes maps based on projects, so under the ‘Projects’ tab you will want to select ‘New project’ to begin building your map. After allowing you to name your project, and with the ‘Default data’ button left checked, selecting the ‘Add’ button will create a basemap of the

world with continents demarcated from oceans and countries outlined lightly in blue on the continents. We will use this as the basis for our map building.

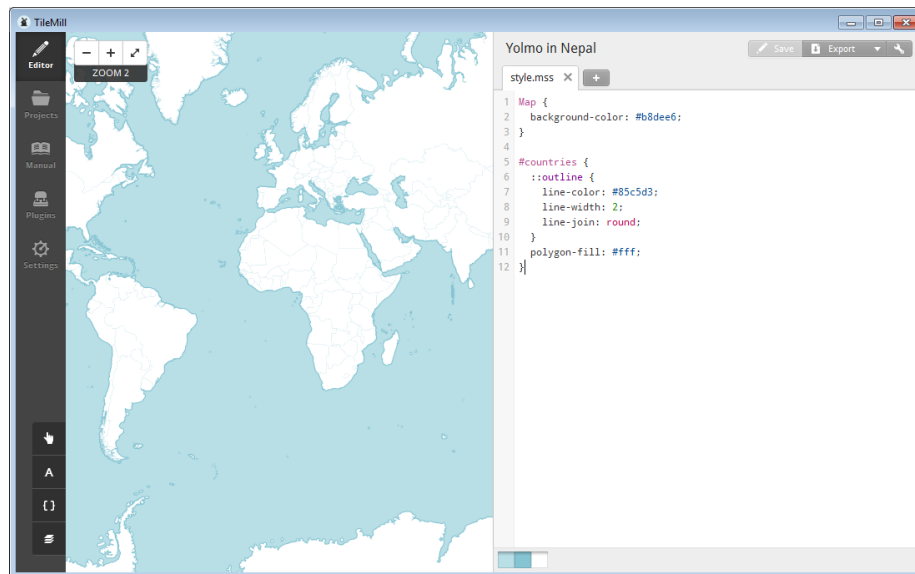


Figure 27. A new TileMill project

On the left side of the screen is the visualization, and on the right is the CartoCSS panel, with the option of adding tabs. We will add another tab when creating labels in order to make accessing the code easier, rather than having it all in one tab. At the bottom of the screen is the color palette being used. As CartoCSS uses hex colors it is useful to have them in one place. The map will only update when you select ‘save’ in the right corner, and we will discuss the export function below once we have built our map. Along the left side, the ‘Projects’ button will take you out to the launch page with all the projects, and the manual is accessible at all times. There is more support further down the left panel. The ‘A’ icon lists the fonts available on the computer, and the ‘{ }’ icon gives the basic CartoCSS codes and a summary of what they do. If you make an error in the syntax of your CartoCSS scripting, TileMill is quite good at indicating specifically where the error is, where possible. The hand icon above these allows for the creation of a basic HTML legend. The lowest icon is where the datasets for building layers are listed, which we will discuss in more detail below.

You can use the zoom to navigate, but unlike CartoDB, TileMill will not automatically update and use that information next time you open the project. Instead, you need to click on the spanner (or wrench) in the top right corner (see Figure 27) to open the project settings and set a default for the project (Figure 28).

Here you can give your project a name and description. You can also limit the zoom range, which will improve performance and project size, as the computer does not have to render as much. Zoom to the preferred level, shift and drag the boundaries and then click to center the map. The centering will then be locked at a specific zoom, as noted in the icon in the Figure above, which is Z7 (Zoom 7), even though we are

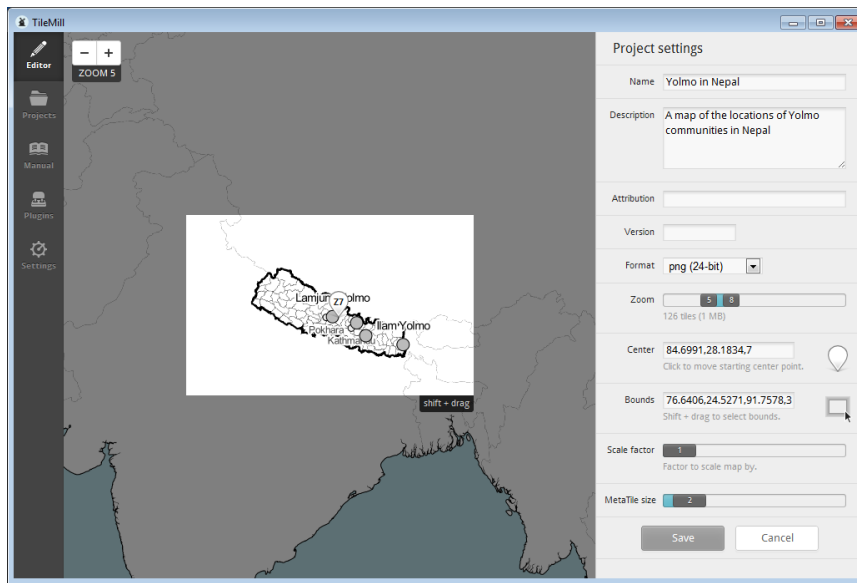


Figure 28. TileMill project settings screen

on zoom level 5. When you are happy with the levels press save, which will exit the screen.

The default TileMill map setup includes the ‘countries’ layer. This gives the ocean and country boundaries. Clicking on the layers button allows you to see this layer and add other datasets for other layers.

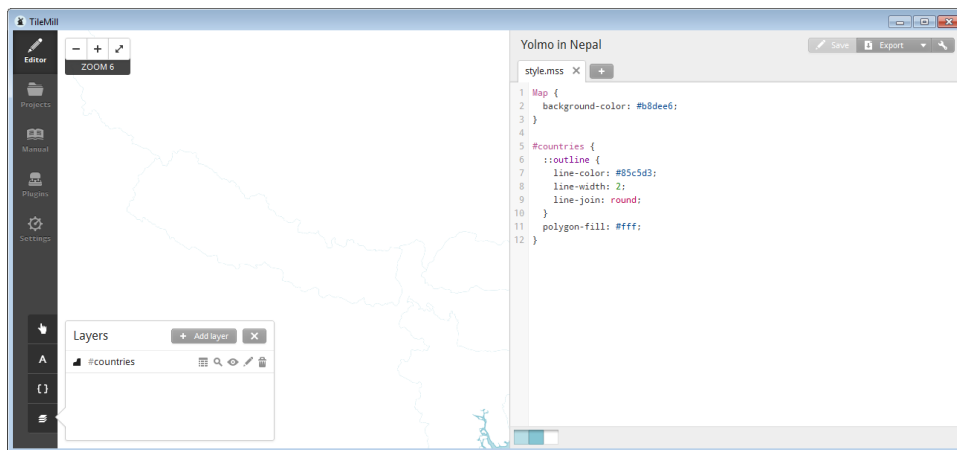


Figure 29. TileMill project layers screen

When you press the ‘Add layer’ button, this will open the screen that allows you to select the next dataset you would like to add. Selecting ‘Browse’ will allow you to navigate to the ‘Mapbox’ server (the middle button on the menu that comes up),

where free GIS data points and layers are stored. If you have map data on your computer (such as those created via Google Maps or Google Earth), navigate to the folder where those files are located, and select the file you wish to add. TileMill supports most map-related files and formats.

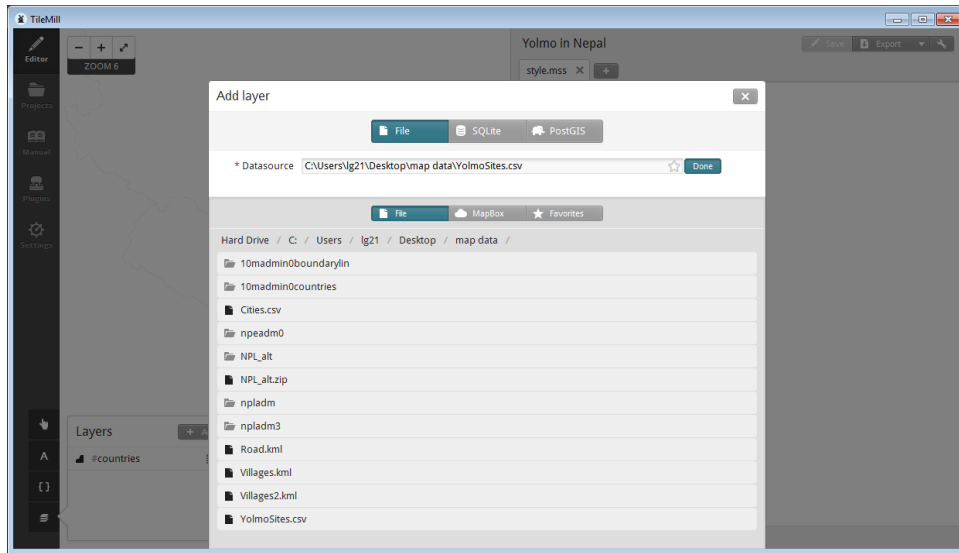


Figure 30. Adding data to TileMill

When you have navigated to your folder and selected the dataset you want to work with, select ‘Save & style.’ By selecting this rather than just ‘save,’ TileMill will create a basic rendering of this data. We selected both the ‘Yolmosites’ and ‘Cities’ data introduced at the start of §3.4.

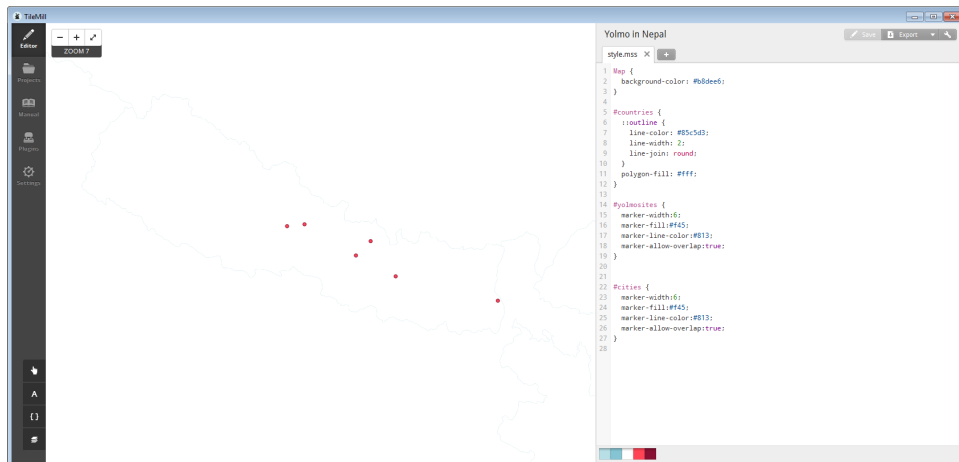


Figure 31. New project with automatically styled points

TileMill has created the points from both datasets with the same size and color. You should notice in Figure 31 that it has also created a small amount of code for each

of the datasets, with the dataset name marked by a hash (e.g., #yolmosites). TileMill uses CartoCSS, a scripting language that can configure color, line width and opacity, labels, and other style features. In order to use these features, some knowledge of the commands and options is necessary. Fortunately, there are a number of tutorials on CartoCSS available online for the interested mapmaker⁵⁴ and the ‘{}’ menu gives definitions of the code used. CartoCSS uses semantically meaningful values. For example, below is the code for the yolmosites points in this default format, which describes how the points should be rendered visually.

```
#yolmosites {
  marker-width:6;
  marker-fill:#f45;
  marker-line-color:#813;
  marker-allow-overlap:true;
}
```

Every time you make a change in the CSS values, press ‘save’ to see the changes. If you are worried you will break something, save the code in a text editor outside TileMill as a backup before making changes. You can change individual values, or you can use the code in Appendix 1, which was used to style the final version of this map. Colors are coded by a series of numbers and letters in hexadecimal, for example, which may require an online tool such as ColorHexa⁵⁵ for you to pick exactly the color you want for a particular element. Unlike in CartoDB, there are no automatically suggested palettes, so you will have to make these design choices for yourself. Figure 32 shows the points above styled so that the #cities data are visually distinct from the #yolmosites data, using the same basic design as in CartoDB in § 3.4 above.

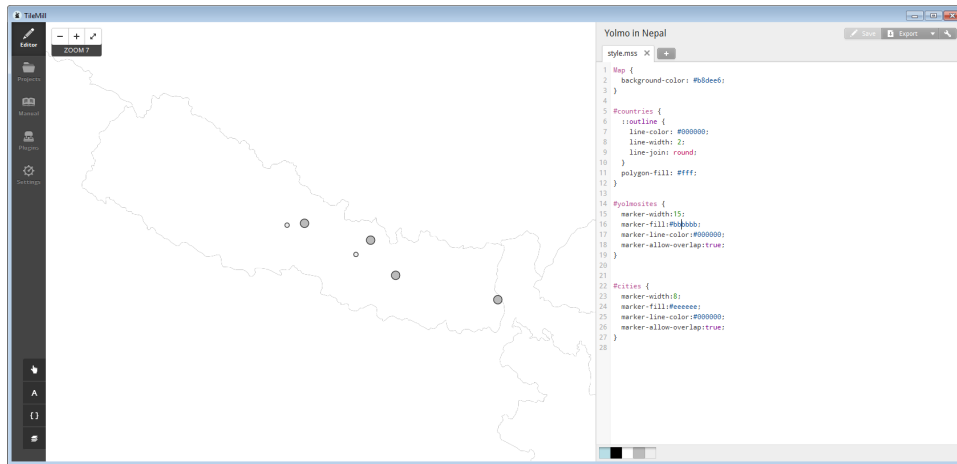


Figure 32. New project with automatically styled points

⁵⁴<http://www.mapbox.com/tilemill/docs/manual/carto/>.

⁵⁵<http://www.colorhexa.com>.

Unlike in CartoDB, you will have to do your own coding to make labels appear. We have started these in a new tab of the coding screen in Figure 33 to help make keeping track of the code a little bit easier. We have once again gone for a style very similar to what we selected in CartoDB. The full code for this map is in Appendix 1. The first line of the label code is a generic font name, then a specific value for that font, and then the generic name marker with the @ is used throughout the code. You can also do the same with colors, specifying a specific hex value be referred to as '@blue' in the code.

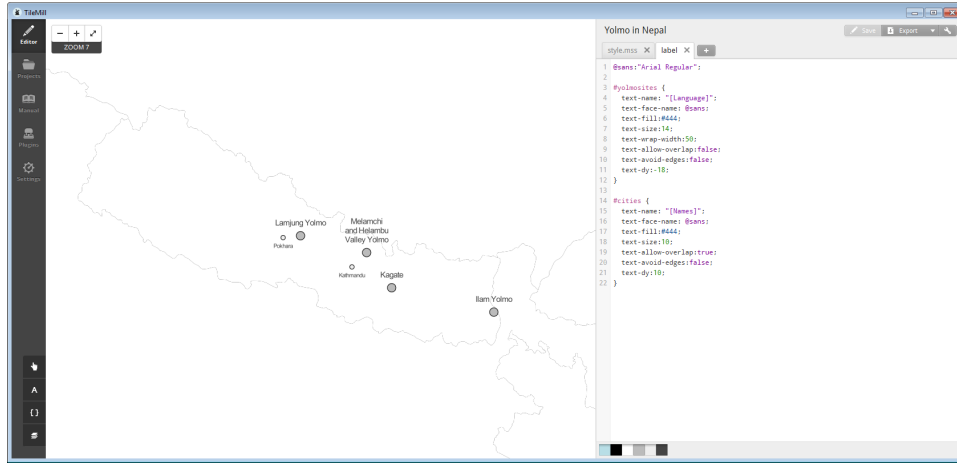


Figure 33. Project with styled points

The map of Nepal is rather underspecified in TileMill in comparison to the detail that comes automatically with a CartoDB basemap. To give some more general locational context we downloaded some Open Access geopolitical boundaries for Nepal from the Global Administrative Areas website.⁵⁶ This data was downloaded as KMZ, which TileMill cannot read, so we opened the files in Google Earth (remember from §3.1 that KMZ is a compressed KML) and resaved them as KML, which TileMill can read. We then chose a style that matched the style of the rest of our map. Although they look like a collection of lines, these are actually polygon data, which means they are solid. For this reason, the order that your layers are listed in becomes important. Placing them lower on the list means they will render underneath the other layers, allowing you to still see the points. This is demonstrated in Figure 34. The polygons do not exactly match TileMill's data, but making the line thicker can help obscure that.

While TileMill allows for a lot more personalization than the other programs we have used, it will still not do everything. You may find that you want to add labels for surrounding countries, or a distance key to a static image of this map. For static maps this is most easily done using a design program on the exported image. The map of Yolmo point data in Figure 3 was generated in TileMill using a similar process

⁵⁶<http://gadm.org/download>. We downloaded their 'Level 0' and 'Level 3' files and renamed them NP00 and NP03 respectively.

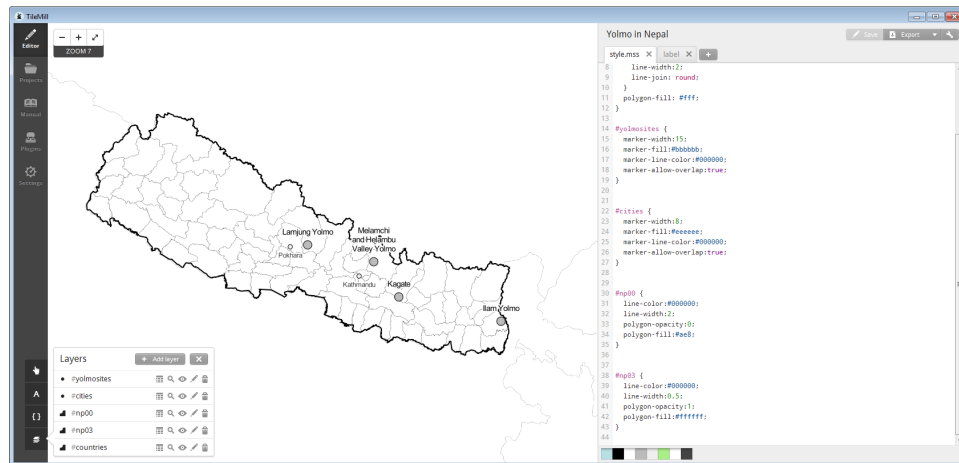


Figure 34. Geopolitical polygon data for Nepal added

to the one demonstrated here. Note that the user had a lot more control over which design elements were included, in comparison to the CartoDB map. Some features, including the India and Tibet labels and the scale, were added after the image was exported from TileMill.

When you have added data and made style adjustments to your layers, TileMill offers an excellent export function. Selecting the ‘Export’ button in the top right of the screen reveals a drop-down menu with options for exporting as PNG, PDF, SVG, and XML. Selecting one of these options opens a new window that defaults to the selection you configured in your project’s settings window. Here you can select new bounds and adjust the actual pixel size of the image to be exported.

There are some issues to be aware of when exporting, as TileMill is primarily designed to make digital maps. ‘Zoom level,’ ‘Scale factor,’ and ‘Metatile size’ will affect how the image and relevant bits such as text will render in your image, and these features interact with the image’s pixel size—increasing the pixel dimensions will result in a higher-quality image but will render text as smaller, for example. This may require multiple exports to finesse your image into a format that works for your purposes. Sometimes, if you have a large enough screen and a small enough map, it can be easier to take a screenshot of the map using your computer’s software.

The clipping of the ‘e’ in Kagate and ‘d’ in Kathmandu in Figure 36 below are because the ‘MetaTile’ level needs to be set higher—go to the general settings panel in the top right corner and increase the size to at least 4. The MetaTile refers to how the software plots and renders map data, in a series of ‘tiles.’ Figure 37 is the same map with MetaTiles increased to 4.

You may want to add more data to your map than you currently have. Adding another layer in TileMill is quite easy, but unfortunately editing such data is not. To add points, lines, or polygons, you will need to create them in another program and import them. To edit this kind of data, you will need to open the underlying layers in

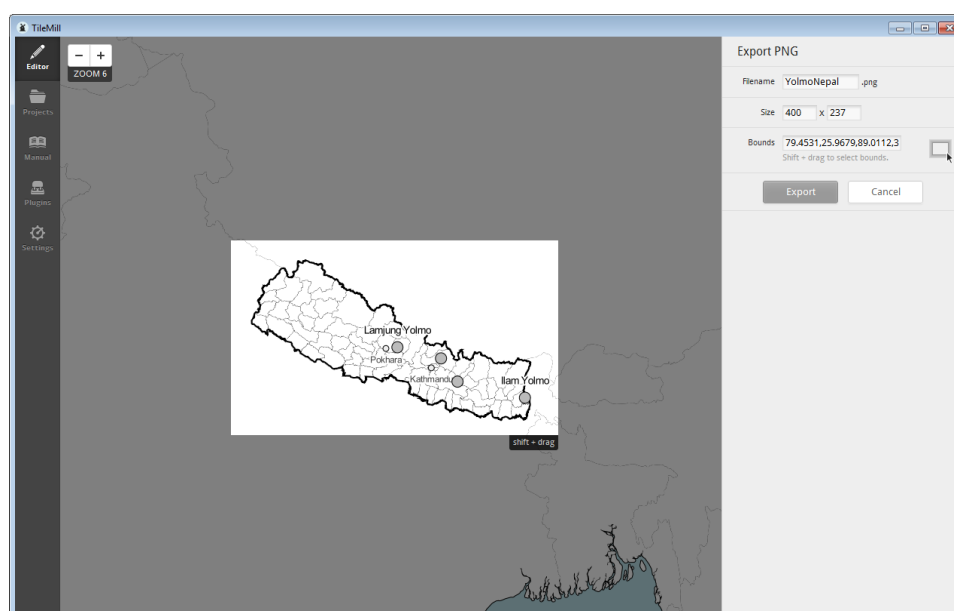


Figure 35. Exporting your map

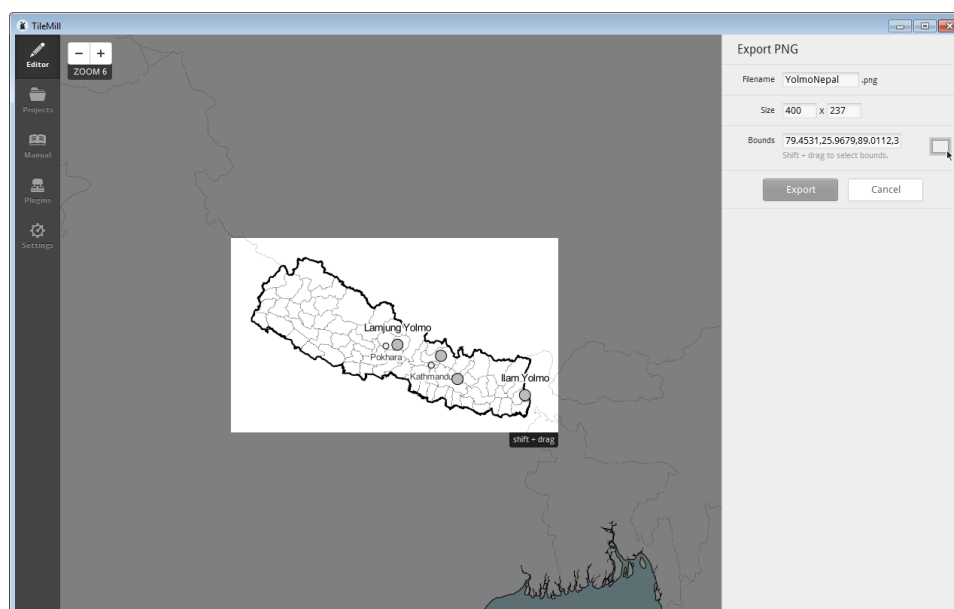


Figure 36. The Yolmo location map built in this tutorial

another program (such as Google Earth), edit them, and re-export them to a format that TileMill can read and render as another layer.

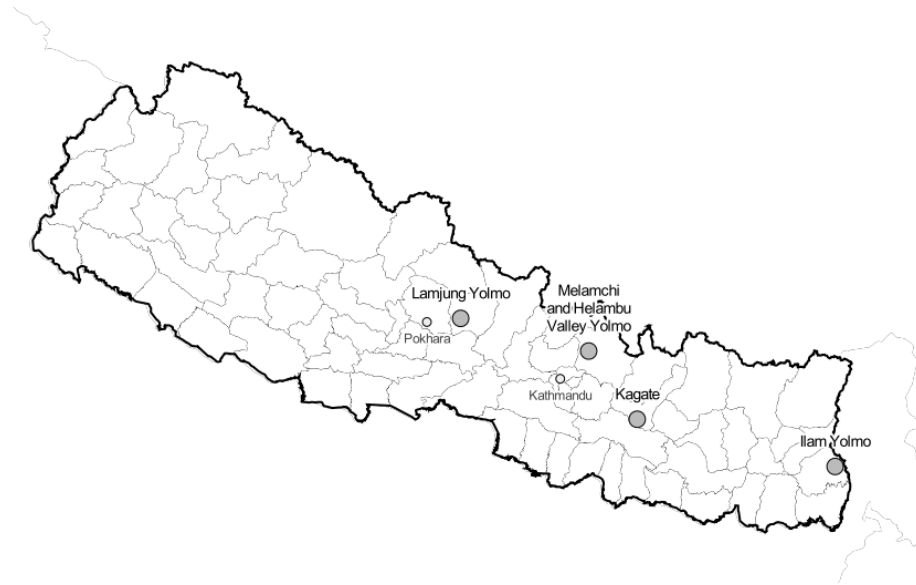


Figure 37. The Yolmo location map built in this tutorial with higher MetaTile count

Finally, we want to discuss two more advanced styling options, which open up a great range of flexibility in designing dynamic maps. The first is styling points in the same dataset in different ways, and the second is styling using zoom levels.

The map we created in this section is a useful illustration of these varieties, but gives the impression that the four varieties of Yolmo are more or less equal. The Melamchi/Helambu area, however, has far more speakers than the other areas, and is the migration source for the other three communities (see Gawne 2013b). Therefore, we may want to give it more prominence. We can do this in CartoCSS by creating some code (below) that will only modify a particular data point. Note that it is exactly the same code as in Appendix 1 for building the basic map. The difference is that a particular point, as indicated by the ‘language’ column header, has been selected in the square brackets at the start of the code, and the marker width has been increased from 15 to 25.

```
#yolmosites[language="Melamchi and Helambu Valley Yolmo"]{
  marker-fill-opacity: 0.9;
  marker-line-color: #000000;
  marker-line-width: 1.5;
  marker-line-opacity: 1;
  marker-placement: point;
  marker-type: ellipse;
  marker-width: 25;
  marker-fill: #aaaaaa;
  marker-allow-overlap: true;
}
```

It is possible to choose all points *except one*, by using ‘!=’ for ‘does not equal’.

```
#yolmosites[language!="Melamchi and Helambu Valley Yolmo"]
  [{
    marker-fill-opacity: 0.9;
    marker-line-color: #000000;
    marker-line-width: 1.5;
    marker-line-opacity: 1;
    marker-placement: point;
    marker-type: ellipse;
    marker-width: 15;
    marker-fill: #aaaaaa;
    marker-allow-overlap: true;
```

Together, these two elements of the code provide the version of the map in Figure 38, with the one variety more prominent than the other three. This kind of styling forms the basis of most categorical or variation mapping.



Figure 38. A variant of the Yolmo map with prominence given to one point

The second advanced styling option is the ability to call up particular features only at particular zoom levels. The specific zoom level is indicated in square brackets in the same way that the specific data type was given above. Appendix 2 gives some code that will work in CartoDB to create the zoom in that active map. This is the start of the #yolmosites code in that appendix:

```
#yolmosites [zoom<11]{
  marker-fill-opacity: 0.9;
```

You can see that this data type is only up to zoom level 11; zooming into level 11 or further will no longer render those points. On the same map, the villages in the Google Earth screenshot in Figure 8 start with this zoom instruction, meaning they will only appear after zoom level 10, putting it in contrast with the other dataset, and so the single point becomes a cluster of smaller village points when zoomed in.

```
#villages [zoom>10]{
  marker-fill-opacity: 0.9;
```

Because the marker width is only relevant to the specific layer, if you are expecting people to zoom in and out, you may want to give a point different values at different levels, such as in this example:

```
#languages ::markers {
  [zoom=9] {marker-width:45}
  [zoom=8] {marker-width:40}
  [zoom=7] {marker-width:35}
  [zoom=6] {marker-width:30}
  [zoom=5] {marker-width:25}
  [zoom=4] {marker-width:20}
  [zoom=3] {marker-width:15}
  [zoom=2] {marker-width:10}
```

To work with this level of zooming complexity is time consuming, and unnecessary if the aim is to make maps for print. It is a good indicator of the kind of complexity that can be coded into a digital map. This kind of zoom functionality can be employed in both TileMill and CartoDB, and requires use of CartoCSS in both cases.

3.6 Sharing your maps Having created beautiful maps, it is understandable that you will also want to share them. In this section we discuss the logistics of distributing static and digital maps, as well as considerations for archiving your data.

Both workflows we described offer options to export maps to static images. These images can be used in traditional publication formats, like journal articles and presentations. Remember to always ensure that you correctly attribute the mapping data, whether it be licensed or whether it is distributable like OpenStreetMap data. Remember to also properly acknowledge the software used to create your language maps to encourage other linguists, researchers and community members to also create their own maps.

Although we have primarily focused on the creation of static maps for display and publication, both workflows also offer the option to share digital maps. Unlike static maps, digital maps allow users to explore, access additional information, and easily disseminate content. When creating your maps in CartoDB and TileMill it is worth considering how you might share your work online as well. With just a small amount more work you may be able to create zoom maps, or maps with additional labels or content for a wider range of audiences. An interactive map on your research

webpage can help people engage with language documentation as a practice situated in real space. Interactive maps can also be incrementally updated and embellished as you gain more data and understanding of an area.

The formats used for storing mapping data make archiving very easy. The KML format is based on XML, which is the same underlying structure as data from programs like ELAN,⁵⁷ and CSV tables are also a suitable format for archiving. Remember to ensure that all points are properly labelled within the file, and sufficient metadata is included to explain what the map data pertains to. Even if you do not create a final map, the data may be archived for future use. High-quality JPEG or PDF versions of your final maps may also be appropriate to archive. Your archivist will be able to inform you of any requirements specific to individual archiving platforms. Remember to also keep backups for your own records. CartoDB is a cloud based platform, so can act as an online backup of your data, but you should download your map whenever it is modified just in case.

As a final consideration, you may wish to make your mapping data more widely available, especially if it is a large areal survey or historical data. For example, Claire Bowerman uses her university website to host KML data of polygons of Australian languages and a map of Pama-Nyungan Languages that she created as part of a project on Australian prehistory.⁵⁸ This data is available for other scholars to use as long as they abide by the conditions set. Alternatively, you could update the information available in resources such as Glottolog, the Endangered Languages Project website, and Wikipedia, or contribute your data towards the OpenStreetMaps initiative. This is particularly useful in countries without much mapping of basic infrastructure. If you have included lines for roads, or points for towns and villages, you can add these as well. As OpenStreetMaps provide much of the basic infrastructure for programs like TileMill and CartoDB, it is a nice way to contribute to open and accessible data.

4. Conclusion Mapping is an extremely useful way to make your data accessible and useful to a larger audience, and this paper only exposes the tip of the iceberg when it comes to making maps. There is an array of resources available both from the creators of Google Earth, CartoDB, and Tilemill, as well as from the mapping community. As with any graphic, a map can represent and contextualise geographical information in a way that a paragraph of text cannot. It can also allow researchers to show correlations between datasets that previously were very difficult to visualize abstractly. Although necessary simplifications and perceived correlations can sometimes be problematic, having such tools in our hands as linguists is an incredible opportunity to communicate specialized knowledge in a clear and succinct way, to fellow researchers and to a wider audience. Along with this opportunity comes the responsibility of presenting the data in an ethical manner and ensuring that the communities we work with are respected. We hope that this brief introduction will help

⁵⁷<http://tla.mpi.nl/tools/tla-tools/elan/>, Max Planck Institute for Psycholinguistics, The Language Archive, Nijmegen, The Netherlands. Sloetjes, Han & Peter Wittenburg 2008.

⁵⁸<http://pamanyungan.sites.yale.edu/language-resources>.

other linguistic researchers to add to the world's knowledge about languages and the communities that speak them through making maps.

References

- Auer, Peter & Jürgen E. Schmidt (eds.). 2010. *Language and space: An international handbook of linguistic variation*. Berlin: Mouton de Gruyter.
- Alfaraz, Gabriela G. 2014. Dialect perceptions in real time: A restudy of Miami-Cuban perceptions. *Journal of Linguistic Geography* 2(2). 74–86. doi: 10.1017/jlg.2014.8
- Bounds, Paulina. 2015. Perceptual regions in Poland: An investigation of Poznan speech perceptions. *Journal of Linguistic Geography* 3(1). 34–45. doi: 10.1017/jlg.2015.1
- Bowern, Claire. 2008. *Linguistic fieldwork: A practical guide*. New York: Palgrave Macmillan.
- Bowern, Claire. 2011. How Many Languages Were Spoken in Australia? *Anggarrgoon: Australian languages on the web*. <http://anggarrgoon.wordpress.com/2011/12/23/how-many-languages-were-spoken-in-australia/>.
- Chamberlain, Brad. 2015. Watersheds and language mapping. *The 25th Annual Meeting of the South East Asian Linguistics Society (SEALS)*. Payap University, Chiang Mai, Thailand, May 27–29.
- Chelliah, Shobhana L., & Willem J. De Reuse. 2011. *Handbook of descriptive linguistic fieldwork*. London: Springer.
- Dahl, Östen, & Ljuba Veselinova. 2005. Language map server. *ESRI User Conference*, July 25–29, San Diego, California.
- Gawne, Lauren. 2013a. Lamjung Yolmo copulas in use: Evidentiality, reported speech and questions. Melbourne: The University of Melbourne dissertation.
- Gawne, Lauren. 2013b. Notes on the relationship between Yolmo and Kagate. *Himalayan Linguistics* 12(2). 1–27.
- Girnth, Heiko. 2010. Mapping language data. In Alfred Lameli, Roland Kehrein & Stefan Rabanus (eds.), *Language and space: An international handbook of linguistic variation 2. Language mapping*, 98–121. Berlin: Mouton de Gruyter.
- Goebl, Hans. 2010. Dialectometry and quantitative mapping. In Alfred Lameli, Roland Kehrein & Stefan Rabanus (eds.), *Language and space: An international handbook of linguistic variation 2. Language mapping*, 433–458. Berlin: Mouton de Gruyter.
- Guier, William H. & George C. Weiffenbach. 1997. Genesis of Satellite Navigation. *John Hopkins APL Technical Digest* 19(1). 178–181.
- Hanewinkel, Christian & Eric Losang. 2010. Animated maps. In Alfred Lameli, Roland Kehrein & Stefan Rabanus (eds.), *Language and space: An international handbook of linguistic variation 2. Language mapping*, 415–433. Berlin: Mouton de Gruyter.
- Hildebrandt, Kristine A., & Shunfu Hu. 2013. Multimedia mapping on the Internet and language documentation: New directions in interdisciplinarity. *Polymath: An Interdisciplinary Arts and Sciences Journal* 3(3). 51–61.

- Hoch, Shawn, & James J. Hayes. 2010. Geolinguistics: The incorporation of geographic information systems and science. *The Geographical Bulletin* 51. 23–36.
- Inoue, Fumio. 2010. Sprachraum and infrastructure: Abstracting geographical space via railway distance. In Alfred Lameli, Roland Kehrein & Stefan Rabanus (eds.), *Language and space: An international handbook of linguistic variation 2. Language mapping*, 542–561. Berlin: Mouton de Gruyter.
- Kehrein, Roland, Alfred Lameli & Stefan Rabanus. 2010. Introduction. In Alfred Lameli, Roland Kehrein & Stefan Rabanus (eds.), *Language Mapping (Language and Space: An International Handbook of Linguistic Variation 2)*, xi–xvii. Berlin: Mouton de Gruyter.
- Kretzschmar, William A., Jr. 1996. Quantitative areal analysis of dialect features. *Language Variation and Change* 8. 13–39.
- Kretzschmar, William A., Jr. 2013a. Computer mapping of Language Data. In Manfred Krug and Julia Schlueter (eds.), *Research methods in language variation and change*, 53–68. Cambridge: Cambridge University Press.
- Kretzschmar, William A., Jr. 2013b. Mapping Southern English. *American Speech* 78. 130–149.
- Kretzschmar, William A., Jr., Ilkka Juuso & C. Thomas Bailey. 2014. Computer simulation for dialect feature diffusion. *Journal of Linguistic Geography* 2(1). 41–57. doi:10.1017/jlg.2014.2
- Krug, Manfred & Julia Schluter (eds.). 2013. *Research Methods in Language Variation and Change*. Cambridge: Cambridge University Press.
- Kulkarni, Vivek, Rami Al-Rfou, Bryan Perozzi & Steven Skiena. 2014. Statistically significant detection of linguistic change. *Computing Research Repository*. doi:abs/1411.3315.
- Labov, William. 2001. *Principles of Linguistic Change, vol. iii: Cognitive and Cultural factors*. Malden and Oxford: Blackwell.
- Labov, William, Sharon Ash, & Charles Boberg. 2006. *The Atlas of North American English*. Berlin: Mouton de Gruyter.
- Labov, William & Dennis R. Preston 2013. Foreword. *Journal of Linguistic Geography* 1(1). 1–3.
- Lameli, Alfred, Roland Kehrein & Stefan Rabanus (eds.). 2010. *Language Mapping. (Language and Space: An International Handbook of Linguistic Variation 2)*. Berlin: Mouton de Gruyter.
- Lewis, M. Paul, Gary F. Simons, & Charles D. Fennig (eds.). 2013. *Ethnologue: Languages of the World*, Seventeenth edn. Dallas, Texas: SIL International. <http://www.ethnologue.com>.
- Luebbering, Candice R. 2013. Displaying the geography of language: The cartography of language maps. *The Linguistics Journal* 7(1). 37–69.
- MacEachren, Alan M. 1995. *How maps work: Representation, visualization, and design*. New York: The Guilford Press.
- Mackey, W.F. 1988. Geolinguistics: Its scope and principles. In C. H. Williams (ed.), *Language in geographic context*, 20–46. Philadelphia: Multilingual Matters.

- MacKenzie, Laurel. 2014. /s/-lenition and the preservation of plurality in modern Occitan. *Journal of Linguistic Geography* 2(2). 59–73.
- Manni, Franz. 2010. Sprachraum and genetics. In Alfred Lameli, Roland Kehrein & Stefan Rabanus (eds.), *Language and space: An international handbook of linguistic variation 2. Language mapping*, 524–542. Berlin: Mouton de Gruyter.
- Montgomery, Chris & Philipp Stoeckle. 2013. Geographic information systems and perceptual dialectology: A method for processing draw-a-map data. *Journal of Linguistic Geography* 1(1). 52–85. doi:10.1017/jlg.2013.4.
- Moore, Samuel, Sanford B. Meech & Harold Whitehall. 1935. *Middle English dialect characteristics and dialect boundaries*. University of Michigan Press: Ann Arbor.
- Mühlhausler, Peter. 2010. Mapping linguistic typology. In Alfred Lameli, Roland Kehrein & Stefan Rabanus (eds.), *Language and space: An international handbook of linguistic variation 2. Language mapping*, 355–375. Berlin: Mouton de Gruyter.
- Murray, Robert W. 2010. Language and space: The Neogrammarian tradition. In Peter Auer & Jürgen E. Schmidt (eds.), *Language and Space: An International Handbook of Linguistic Variation*, 70–87. Berlin: Mouton de Gruyter.
- Ormeling, Ferjan. 2010. Visualizing geographic space: The nature of maps. In Alfred Lameli, Roland Kehrein & Stefan Rabanus (eds.), *Language and space: An international handbook of linguistic variation 2. Language mapping*, 21–40. Berlin: Mouton de Gruyter.
- Parkinson, Bradford W. 1996. Introduction and heritage of NAVSTAR, the Global Positioning System. In Bradford W. Parkinson, James J. Spilker, Penina Axelrad & Per Enge (eds.), *Global Positioning System theory and applications*, vol. 1, 3–28. Washington, D.C.: American Institute of Aeronautics and Astronautics. <http://site.ebrary.com/id/10516767>.
- Peeters, Y. J. D. 1992. The political importance of the visualisation of language contact. *Discussion Papers in Geolinguistics* 19–21. 6–8.
- Pellerin, Cheryl. 2006. United States updates Global Positioning System technology: New GPS satellite ushers in a range of future improvements. IIP Digital. US Embassy. <http://iipdigital.usembassy.gov/st/english/article/2006/02/20060203125928lc-nirellepo.5061609.html#ixzz3cANu1lpE>.
- Preston, Dennis R. 2010. Mapping the geolinguistic spaces of the brain. In Alfred Lameli, Roland Kehrein & Stefan Rabanus (eds.), *Language and space: An international handbook of linguistic variation 2. Language mapping*, 121–141. Berlin: Mouton de Gruyter.
- Prichard, Hilary. 2014. Northern dialect evidence for the chronology of the Great Vowel Shift. *Journal of Linguistic Geography* 2(1). 87–102. doi:10.1017/jlg.2014.9.
- Roth, Tim. 2013. The major dialects of Nyamwezi and their relationship to Sukuma: A time-based perspective. *Journal of Linguistic Geography* 1(2). 125–140. Cambridge University Press. doi:10.1017/jlg.2013.11
- Scheuringer, Hermann. 2010. Mapping the German language. In Alfred Lameli, Roland Kehrein & Stefan Rabanus (eds.), *Language and space: An international*

- handbook of linguistic variation 2. Language mapping*, 158–180. Berlin: Mouton de Gruyter.
- Sloetjes, Han & Peter Wittenburg. 2008. Annotation by category – ELAN and ISO DCR. *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008)*, 816–820.
- Stanford, James. 2009. Eating the food of our place: Sociolinguistic loyalties in multi-dialectal Sui villages. *Language in Society* 38(3). 287–309.
- Swiggers, Pierre. 2010. Mapping the Romance languages of Europe. In Alfred Lameli, Roland Kehrein & Stefan Rabanus (eds.), *Language and space: An international handbook of linguistic variation 2. Language mapping*, 269–301. Berlin: Mouton de Gruyter.
- Trudgill, Peter. 1983. *On dialect: Social and geographical perspectives*. Oxford: Blackwell.
- Upton, Clive. 2010. Designing maps for non-linguists. In Alfred Lameli, Roland Kehrein & Stefan Rabanus (eds.), *Language Mapping (Language and Space: An International Handbook of Linguistic Variation 2)*, 142–156. Berlin: Mouton de Gruyter.
- Uytvanck, Dieter van, Alex Dukers, Jacqueliijn Ringersma & Paul Trilsbeek. 2008. Language-Sites: Accessing and presenting language resources via geographic information systems. In N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odiijk, S. Piperidis & D. Tapias (Eds.), *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008)*, Marrakech, Morocco. European Language Resources Association (ELRA).
- Waller, Alessia Valentini. 2009. “*Il Toscano non è un dialetto*”: *Variation in Italian language attitudes*. Atlanta, Georgia: Emory University thesis. <http://pid.emory.edu/ark:/25593/1b5tj>.
- Wattel, Evert & Pieter van Reenen. 2010. Probabilistic maps. In Alfred Lameli, Roland Kehrein & Stefan Rabanus (eds.), *Language and space: An international handbook of linguistic variation 2. Language mapping*, 495–504. Berlin: Mouton de Gruyter.
- Williams, Colin H. 1988. An introduction to geolinguistics. In Colin H. Williams (ed.), *Language in geographic context*, 1–19. Philadelphia: Multilingual Matters.
- Williams, Colin H., & Ambrose, J.E. 1988. On measuring language border areas. In C.H. Williams (ed.), *Language in geographic context*, 93–135. Philadelphia: Multilingual Matters.
- Yan, Qingyang. 2015. The perceptual categorization of Enshi Mandarin regional varieties. *Journal of Linguistic Geography* 3(1). 1–19. doi: 10.1017/jlg.2015.3.

Appendix 1 Below is the CSS code used to build the Yolmo map using TileMill in §3.5. Many of these design elements were also employed in the CartoDB map in §3.4 as well. The code can be downloaded from <https://scholarspace.manoa.hawaii.edu/bitstream/10125/24692/4/appendix1.css>.

```
Map {
  background-color: #b8dee6;
}

#countries {
  ::outline {
    line-color: #000000;
    line-width: 2;
    line-join: round;
  }
  polygon-fill: #fff;
}

#yolmosites {
  marker-width: 15;
  marker-fill: #bbbbbb;
  marker-line-color: #000000;
  marker-allow-overlap: true;
}

#cities {
  marker-width: 8;
  marker-fill: #eeeeee;
  marker-line-color: #000000;
  marker-allow-overlap: true;
}

#np00 {
  line-color: #000000;
  line-width: 2;
  polygon-opacity: 0;
  polygon-fill: #ae8;
}

#np03 {
  line-color: #000000;
  line-width: 0.5;
}
```

```

    polygon-opacity:1;

    polygon-fill:#ffffff;
}

@sans:"Arial Regular","Liberation Sans Regular","DejaVu
    Sans Book";

#yolmosites {
    text-name: "[Language]";
    text-face-name: @sans;
    text-fill:#000000;
    text-halo-fill:#ffffff;
    text-halo-radius:2;
    text-size:14;
    text-wrap-width:50;
    text-allow-overlap:false;
    text-avoid-edges:false;
    text-dy:-18;
}

#cities {
    text-name: "[Names]";
    text-face-name: @sans;
    text-fill:#444;
    text-halo-fill:#ffffff;
    text-halo-radius:2;
    text-size:12;
    text-allow-overlap:true;
    text-avoid-edges:false;
    text-dy:10;
}

```

Appendix 2 Below is the CSS code used to make the interactive Yolmo map, which can be found here:

http://lgawne.cartodb.com/viz/94f75f72-a200-11e4-90b6-0e0c41326911/public_map.

#villages refers to Yolmo villages and #villages2 refers to non-Yolmo villages. The #yolmosites data is the same as in Appendix 1, but because of the zoom restrictions will not be visible from zoom level 11. The code can be downloaded from <https://scholarspace.manoa.hawaii.edu/bitstream/10125/24692/5/appendix2.css>.

```

#yolmosites [zoom<11]{
    marker-fill-opacity: 0.9;
    marker-line-color: #000000;

```

```
marker-line-width: 2;
marker-line-opacity: 1;

marker-placement: point;
marker-type: ellipse;
marker-width: 20;
marker-fill: #D6301D;
marker-allow-overlap: true;
}

#yolmosites::labels [zoom<11]{
text-name: [language];
text-face-name: 'Lato Regular';
text-size: 15;
text-label-position-tolerance: 0;
text-fill: #FFFFFF;
text-halo-fill: #000000;
text-halo-radius: 2;
text-dy: -20;
text-allow-overlap: true;
text-placement: point;
text-placement-type: dummy;
}

#villages[zoom>10]{
marker-fill-opacity: 0.9;
marker-line-color: #FFF;
marker-line-width: 2;
marker-line-opacity: 1;
marker-placement: point;
marker-type: ellipse;
[zoom>10] {marker-width:15}
marker-fill: #D6301D;
marker-allow-overlap: true;
}

#villages::labels [zoom>10]{
text-name: [name];
text-face-name: 'Lato Regular';
text-size: 14;
text-label-position-tolerance: 0;
text-fill: #000;
text-halo-fill: #FFFFFF;
text-halo-radius: 2;
text-dy: -15;
```

```
text-allow-overlap: true;
text-placement: point;
text-placement-type: dummy;
}

#villages2[zoom>10]{
  marker-fill-opacity: 0.9;
  marker-line-color: #FFF;
  marker-line-width: 1.5;
  marker-line-opacity: 1;
  marker-placement: point;
  marker-type: ellipse;
  marker-width: 12;
  marker-fill: #3E7BB6;
  marker-allow-overlap: true;
}

#villages2::labels [zoom>10]{
  text-name: [name];
  text-face-name: 'Lato Regular';
  text-size: 14;
  text-label-position-tolerance: 0;
  text-fill: #000;
  text-halo-fill: #FFF;
  text-halo-radius: 1.5;
  text-dy: -12;
  text-allow-overlap: true;
  text-placement: point;
  text-placement-type: dummy;
}

#road[zoom>10]{
  line-color: #FFCC00;
  line-width: 2;
  line-opacity: 0.7;
}
```

Lauren Gawne
lg21@soas.ac.uk

Hiram Ring
hiram1@ntu.edu.sg